Addressing unanswered questions in bacterial hydrocarbon biosynthesis

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Matthew Ryan Jensen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Carrie M. Wilmot, Advisor

October, 2017

Acknowledgements

Carrie: Thank you, Carrie, for all of mentorship, guidance, and patience throughout graduate school. I am eternally grateful for the opportunity to learn and grow as a scientist under your leadership. Thank you for the chance to make my grad school home in your lab. I have big shoes to fill, but I will dedicate my career to living up to the expectations and potential you saw in me.

Kendra: You are my best friend and truly my better half. Thank you for encouraging me to move to the Twin Cities for graduate school even though it meant living apart for 3.5 years. On that note, thank you, Skype, for allowing us to talk and hang out every night from August 2012-May 2015. Kendra, you have been by rock throughout these last five years, and I could not have asked for a better partner in life. Thank you for always encouraging me and challenging me to be my best self. I couldn't have done any of this without you. Love you!

Family: Thank you all for your love and support during this journey. Growing up and moving far from home sucks, so thank you all for being there for me. I really appreciate your advice, encouragement, jokes, and all around silliness. I love visiting you all back in Nebraska, so thank you for making my trips back home a great destination. Your frequent trips up north have really made my time here fun and exciting. Sibling threads on Facebook and Snapchat (The Big Kids Table) have helped me laugh and stay connected as if I never left home. I love you all. I have the best parents and siblings in the world, and for that, I am truly blessed.

Friends: Thank you to all of the wonderful friends I have made during graduate school. You have been my source of fun, inspiration, aid, support, and companionship, and for that I am truly grateful. You have made these past five years an adventure full of meals, study breaks, talks, laughs, concerts, happy hours, trips, weddings, 6-pack challenges, solar eclipses, coffee roasting, cabins, Itasca weekends, bonfires, fantasy football, tubing down the Zumbro River, and all around the most fun I've had in my life. You all make balancing work and life easy and enjoyable. I want to thank my lab mates, Bob, Chao, and Morgan for being wonderful friends and colleagues, as well as fostering an amazing lab environment. I also want to thank the Wackett Lab for literally everything, with special thanks to the lunch gang: James, Kelly, Adi, Diego, and many more.

U of MN: I would like to thank my dissertation committee members, Doug Ohlendorf (chair), Jeff Gralnick, Mark Distefano, and Carrie Wilmot (PI/Advisor). I want to thank the graduate school, CBS, BMBB (Sarah and Sue), and BTI (Lisa and everyone in the office). I would also like to thank everyone involved with the NIH Chemistry-Biology Interface Training Grant through which wonderful opportunities were afforded me.

Dedication

This dissertation is dedicated to my wife, friends, family, and colleagues.

Abstract

Modern society relies heavily on hydrocarbons. Because they are used as liquid transportation fuels, cosmetics, waxes, and food coatings, hydrocarbons are important components of most aspects of daily life. The majority of hydrocarbon products are extracted or derived from cude oil. Energy costs, fuel demand, and environmental concerns involving the non-renewable nature of petroleum-derived compounds have sparked recent interest in microbial hydrocarbon production. Engineering the diversity of microbial metabolic pathways to produce biofuels and chemicals represents a renewable alternative to fossil fuels. One pathway of interest is bacterial long-chain olefin biosynthesis. Divergent bacterial species have been shown to synthesize these waxy hydrocarbons using four enzymes: OleABCD. Recent investigations have aimed to understand how these enzymes work in concert to produce valuable hydrocarbon intermediates and products. These findings will be useful for future pathway engineering for renewable, bacterial olefin production.

The first three chapters of this dissertation deal with elucidating the catalytic mechanism of the first enzyme in the olefin biosynthesis pathway, OleA. Chapters 2, 3, and 4 each investigate a separate amino acid necessary for OleA β-keto acid formation using methods of site-directed mutagenesis, biochemistry, and X-ray crystallography. In Chapter 2, the unique substrate binding channel architecture of OleA is directly demonstrated by trapping substrates and intermediates within Cys143 mutated enzymes. The role of Glu117 as the

catalytic base needed to prime condensation through deprotonation of the second acyl-CoA substrate is established in Chapter 3. This represents the first dimeric thiolase superfamily enzyme that uses an active site base donated from the second monomer. It also provides evidence for the unique mechanistic strategy of OleA compared to other thiolases. Chapter 4 investigates the role His285 plays in positioning substrate and intermediates for productive condensation by OleA. It is also shown that His285 plays a role in protecting the Cys143 thiolate from oxidative damage.

The dissertation concludes with the investigation of the catalytic function of OleC and the characterization of the multienzyme assembly formed by OleB, OleC, and OleD. In Chapter 6, OleC is demonstrated to produce β-lactones from β-hydroxy acids. This is the first example of a β-lactone synthetase, a novel enzyme function. It is also shown that OleC is homologous to amino acid sequences encoded in known β-lactone-producing natural product pathways, suggesting a common mechanism for β-lactone formation. Chapter 7 details the formation of an assembly consisting of OleBCD. Following co-expression of OleABCD, OleBCD are found to co-elute over nickel-affinity, anti-FLAG, and size-exclusion chromatographic purifications. These assemblies form ~2 MDa structures that produce *cis*-olefin following the addition of OleA and acyl-CoA. Negative stain transmission electron microscopy reveals a mixture of assemblies ranging from 24-40 nm in diameter. It is proposed that these assemblies are necessary for protecting the cell from the highly-reactive β-lactone intermediate.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Replacing fossil fuels with renewable hydrocarbons

Modern society relies heavily upon hydrocarbons found in nature. These chemicals are widespread for their uses as fuels, lubricants, polymers, and wax coatings, to name a few. The most widely used hydrocarbon products, such as liquid petroleum-based fuels, derive from non-renewable crude oil, the source accounting for greater than 80% of our energy needs (1). Global consumption of crude oil has risen to nearly 35 billion barrels per year (International Energy Agency, 2016). The non-renewable nature of fossil fuels coupled with increasing consumption rates and environmental concerns necessitates the transition from reliance on fossil fuel-derived hydrocarbons to renewable hydrocarbons. Even though renewable energy (solar, wind, hydroelectric, etc.) has made promising advances in reducing dependence on crude oil, these energy forms do not negate the need for energy-dense liquid fuels for the transportation industry.

Early attempts to reduce dependence on petroleum were focused on plant mass fermentation to produce biofuels. First generation biofuels, such as ethanol and plant oil-derived biodiesel, were attractive targets due to their renewable nature and availability. However, both biofuels presented significant problems. Bioethanol is hygroscopic and contains only 70% of the energy per unit mass of

highly-reduced hydrocarbons found in crude oil (2). Biodiesel from plant oils, as well as ethanol, negatively impacts the food industry by competing for land use and crop yields (3). To skirt these potential pitfalls, researchers turned to microbial biosynthesis pathways to produce gasoline- and diesel-length hydrocarbons. Microbes have been known to produce an array of important hydrocarbons, such as alkanes, alkenes, arenes, alcohols, and isoprenoid compounds (4). As these pathways produce hydrocarbons from central metabolic pathways, production by fermentation of sugars from pretreated biomass represents an attractive target (Figure 1A). Ideally, researchers sought a platform of microbial fermentation that secretes a water immiscible product that is compatible with current petroleum refinery operations (Figure 1B) (4). However, many cost-prohibitive problems with efficiently preparing biomass feedstocks remain, such as chemical pretreatments and enzymatic degradation of lignocellulosic material to convert biomass to fermentable sugars (5).

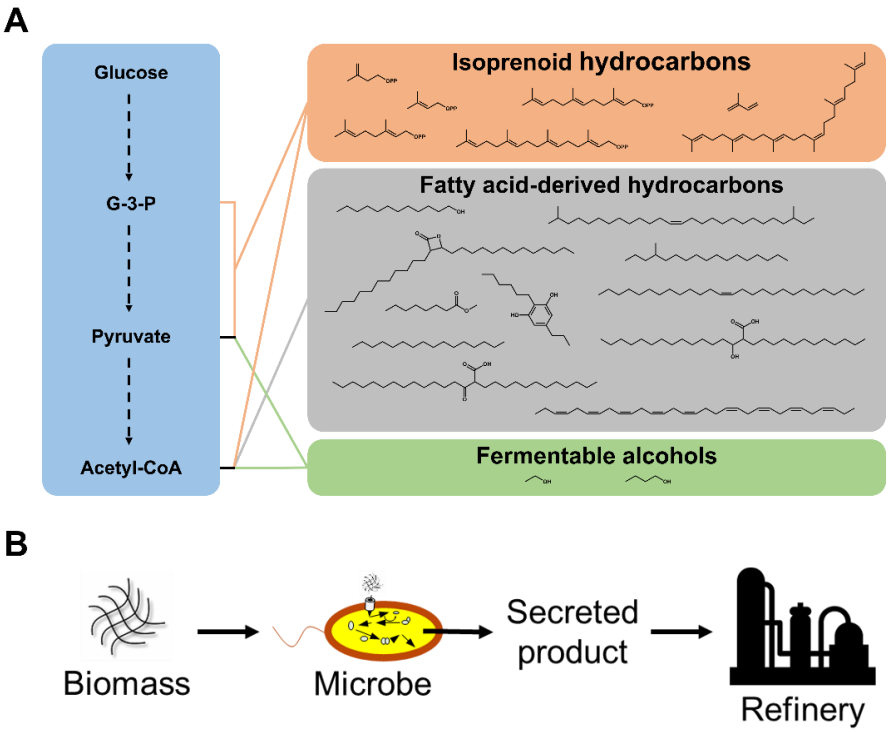


Figure 1: Microbial pathways for renewable hydrocarbon production. **A)** Major categories of microbial biofuel hydrocarbons and the pathways from which they derive. Major examples of chemical structures are provided for each category. **B)** Ideal platform for fermentation of pretreated biomass by specialty microbes, producing secreted hydrocarbons that are compatible with current refinery operations. Panel A adapted from (4). G-3-P, glyceraldehyde-3-phosphate.

Apart from replacement biofuels, the pursuit of microbial hydrocarbons for other important types of compounds represents a viable path for reducing our dependence on oil. In a single barrel of crude oil, ~75% of the volume yields hydrocarbons of the length and makeup for liquid transportation fuels (U.S. Energy Information Administration, 2014) (Figure 2). As they are the major products from crude oil, biofuels are an obvious area to research. However, the demand for liquid fuels in the transportation sector is far too high to keep up with renewable production methods at the present time. For example, ocean tankers

consume more than one US ton of liquid fuel for every three miles traveled (6). A more commercially practical target in the short term is the 17% of hydrocarbon volume of a barrel of oil, which is used across many industries as lubricants, waxes, water sealants, chemical synthons, polymer building blocks, solvents, personal care products, cosmetics, and candy food coatings (Figure 2). Microbial production of these types of hydrocarbons could more realistically meet market demands than replacing the heavy demand of liquid transportation fuels. Specific hydrocarbons can be produced by and extracted from microbes as opposed to refining and purifying individual or groups of compounds from crude oil. Microbial synthesis also avoids environmental concerns of alternative production methods such as extracting waxes and other long-chain hydrocarbons from endangered species. One example of attainable and profitable microbial production of these types of compounds is from the company Amyris. Their renewable production of squalene, a hydrocarbon used in cosmetics, accounts for nearly 25% of the $93 million squalene industry (ETC Group, 2014). A greater understanding of the cellular pathways in bacteria, yeast, algae and other microbes that produce these types of compounds or precursor compounds is an important and necessary step toward renewable hydrocarbon production.

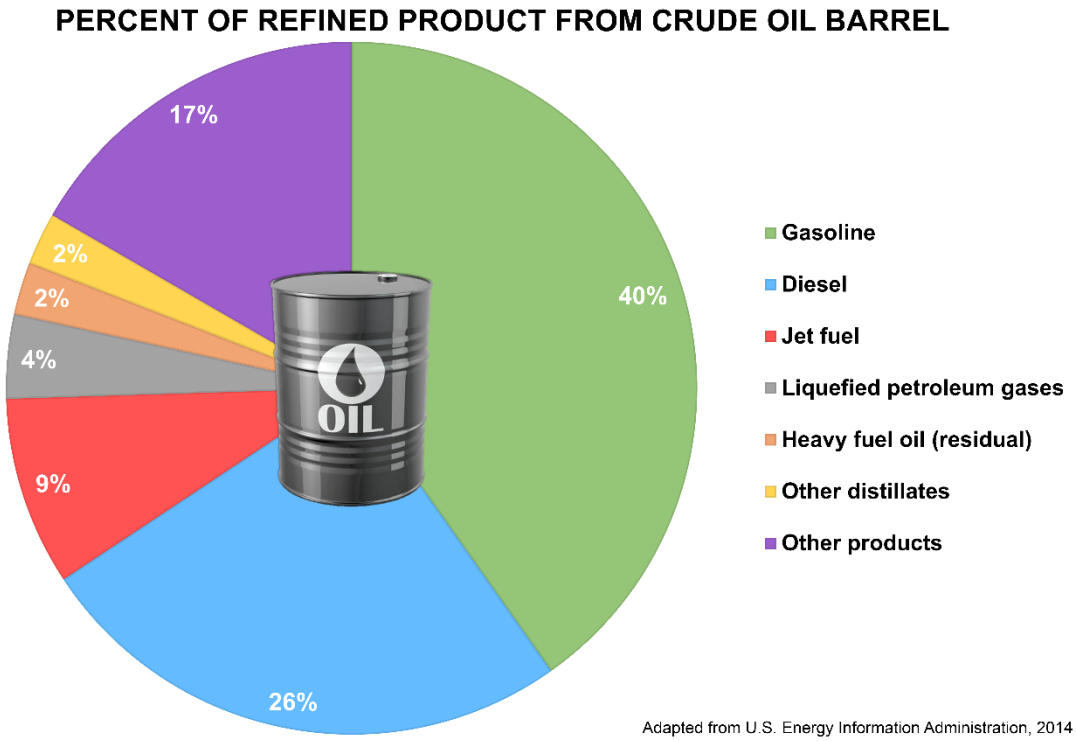


Figure 2: Breakdown of the refined products from a single barrel of crude oil. Gasoline (green), diesel (blue), and jet fuel (red) are considered liquid transportation fuels. The other products category (purple) represents a financially viable option for meeting current demands of commodity hydrocarbons by renewable microbial production. Percentage values taken from the U.S. Energy Information Administration, 2014.

Long-chain olefin biosynthesis pathway

The bacterial biosynthesis pathway for long-chain olefins is in the subject of this dissertation. Long-chain (C_{23}-C_{33}) olefins consist of two hydrocarbon chains connected by a *cis*-double bond at the median position. They differ in chain length, unsaturation of alkyl chains, and various modifications such as methyl branching (7). These hydrocarbons, as well as intermediates during their biosynthesis, represent a renewable class of molecules to replace petroleum-based commodity chemicals, waxes, and lubricants.

Long-chain olefins have been studied for over 50 years. In 1969, researchers described these hydrocarbons as forming their characteristic double bond via a “head-to-head” condensation of two fatty acids at their carboxylic acid “head” groups (8). Nearly 40 years later the genes and enzymes necessary for olefin production, *oleABCD*, were identified and patented by the company LS9, Inc. (now Renewable Energy Group) (9,10). Several reports followed this patent describing the three- and four-gene clusters found in phylogenetically diverse bacteria. Frias and coworkers showed that the dominant stereochemical product made by this pathway in *Arthrobacter* species is a *cis*-olefin (11). In 2010, Beller and coworkers described the three-gene cluster in *Micrococcus luteus* comprised of *oleA, oleD,* and a fusion of *oleB* and *oleC*, proposing a mechanism for alkene formation by these and other gene products (12). In the same year Sukovich and coworkers described the four-gene cluster found in *Shewanella oneidensis* as well as 68 other bacteria, placing each gene product into a known enzyme superfamily to suggest their function (7,13). Gas chromatography-mass spectrometry (GC/MS) analyses of the membrane composition from 14 of these bacteria illustrated the diversity of hydrocarbons produced by species-specific variants of this pathway (13). *Xanthomonas campestris* was shown to produce at least 15 different olefin products containing one, two, and three double bonds. Other species, such as the cold water bacterium *S. oneidensis*, produced a single polyolefin, 3,6,9,12,15,19,22,25,28-hentriacontanonaene. Strains producing this polyolefin exhibited a shorter lag-phase when growth was shifted

to colder temperatures compared to strains lacking *oleABCD* (7). It is hypothesized that olefinic hydrocarbons produced by the olefin biosynthesis pathway alter membrane fluidity; however, more evidence is needed to understand why olefins are produced by these bacteria (7).

OleA

The first step of the olefin biosynthesis pathway is catalyzed by OleA. OleA is a member of the thiolase superfamily of enzymes, which are dimeric enzymes that catalyze condensation reactions with an array of activated fatty acid substrates (14). Like all thiolase enzymes, OleA contains a cysteine nucleophile that forms a transient thioester-acyl intermediate during the catalytic cycle. In the context of olefin biosynthesis, OleA condenses two long-chain fatty acyl-coenzyme A (CoA) substrates to form a β-keto acid (Figure 3) (15). It was demonstrated through heterologous expression studies that OleA directs the type of olefin ultimately produced by controlling which fatty acyl-CoA substrates are condensed in the first step (13). *X. campestris* OleA was shown to condense a range of fatty acyl-CoA substrates varying in chain length and saturation, consistent with the ability of this organism to produce a wide variety of olefin products (Table 1) (13,15).

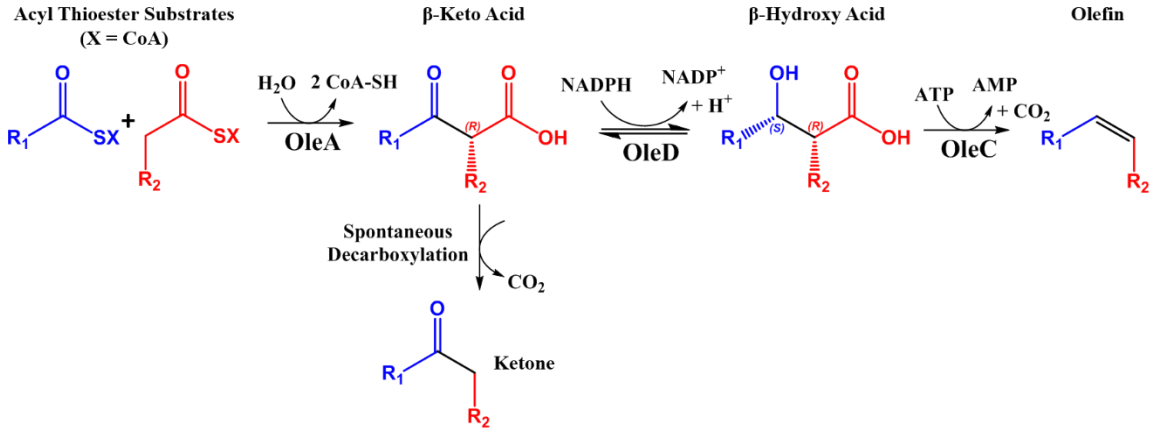


Figure 3: The first proposed mechanism of long-chain olefin biosynthesis. OleA catalyzes the condensation of two acyl-CoA molecules to form a β-keto acid. OleD reduces this product to a (2*R*,3*S*)-β-hydroxy acid. OleC is hypothesized to carry out the final decarboxylation and dehydration required to form olefin. This scheme does not define a role for OleB. Figure adapted from (16).

Table 1: Substrate specificity of *X. campestris* OleA as determined by CoA-SH release. (15)

Substrate			
Common name	**Carbon chain length**	**CoA-SH product (μM)[a]**	**Percent yield (%)[b]**
Palmitoyl-CoA	16	65.0 ± 0.9	100
Myristoyl-CoA	14	63.2 ± 0.4	97
Lauroyl-CoA	12	51.4 ± 1.9	79
Palmitoleoyl-CoA	16	36.9 ± 0.9	57
Decanoyl-CoA	10	27.2 ± 1.6	42
Stearoyl-CoA	18	18.7 ± 1.8	29
Octanoyl-CoA	8	8.0 ± 2.2	12
Acetyl-CoA	2	ND[C]	ND

Values shown are the average in triplicate with standard error.
[a] Free coenzyme A detected by absorption at 412 nm following reaction with 5,5'-dithio-bis-(2-nitrobenzoic acid).
[b] Starting substrate was 65 μM; 65 μM product is 100% theoretical yield.
[c]ND, no detectable activity.

OleD

OleD catalyzes the second step of olefin biosynthesis. Predicted as a member of the short-chain dehydrogenase/reductase superfamily, OleD from *Stenotrophomonas maltophilia* was purified and characterized in 2011 by Bonnett

and coworkers (13,16). In this study OleD was shown to be a dimeric enzyme that required NADPH for activity (16). Synthetic substrates of variable chain length and stereochemistry were reacted with OleD in the on-pathway reductive direction as well as the oxidative direction. It was shown that OleD exhibits strong stereochemical preference for (2*R*,3*S*)-β-hydroxy acids in the oxidative reaction, indicating that the true product, at least for *S. maltophilia*, likely contains this stereochemical configuration (16). Reactions with (2*R/S*)-β-keto acids in the reductive direction demonstrated that OleD accepts the β-keto acid produced by OleA, further supporting the findings of Frias and coworkers (15). It was also implied that the condensation step of OleA is responsible for introducing the proper 2*R* stereochemistry at C_2 of the β-keto acid. Together, these results showed the stereochemical course for the first two steps of olefin biosynthesis, where OleA produces a (2*R*)-β-keto acid that is reduced to the (2*R*,3*S*)-β-hydroxy acid by OleD (Figure 3) (16).

OleC

OleC is a member of the acyl-CoA synthetase-nonribosomal peptide synthetase adenylation domain-luciferase (ANL) superfamily. Enzymes of this superfamily generally catalyze two partial reactions, the first being an adenylation of the substrate carboxylate to form an acyl-AMP intermediate (17). OleC from *S. maltophilia* was first expressed and purified in 2010 by Frias, Goblirsch, and coworkers (18). In this study, OleC was crystallized in the presence of AMP, and X-ray diffraction data were collected for this crystal. However, due to low

resolution data (3.4 Å) and poor amino acid sequence homology with other ANL superfamily members, molecular replacement was unsuccessful in phasing the dataset. No further structural investigations of OleC have been published to date.

The ability of OleC to accept β-hydroxy acid substrates of varying stereochemical configurations was described by Kancharla and coworkers in 2016 (19). They found that OleC from *S. maltophilia* was able to convert all four stereoisomers of 2-hexyl-3-hydroxydecanoic acid to *cis*-olefin via an ATP-dependent decarboxylative dehydration (19). It was posited that the adenylated intermediate is formed by reaction of the C_3 hydroxyl with ATP to facilitate decarboxylation (Figure 4) (19). This hypothesis would make OleC mechanistically different from other ANL superfamily enzymes. These findings led to the assignment of OleC catalyzing the final step of olefin biosynthesis, marking OleACD as sufficient for olefin production. However, our recent work with *X. campestris* OleC has demonstrated that the initial assignment of olefin as the OleC product was artefactual, and the true product of OleC catalysis is revealed in Chapter 6 of this dissertation.

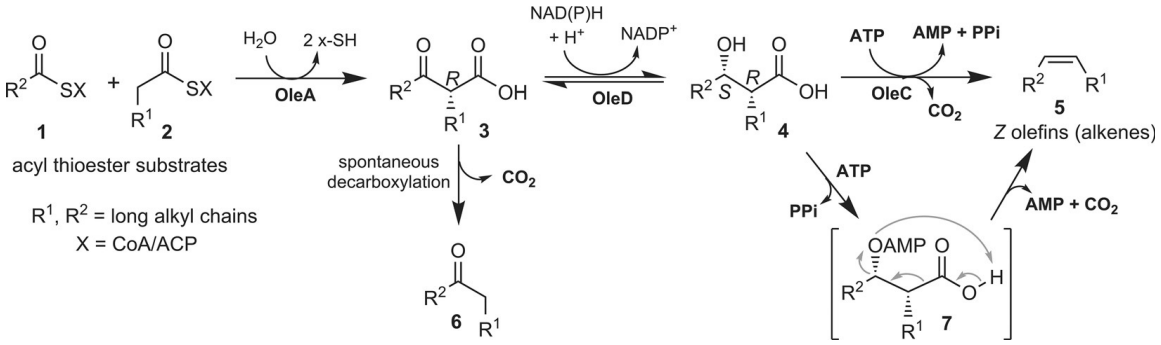


Figure 4: Mechanism for olefin biosynthesis as proposed by Kancharla and coworkers (19). This mechanism lacks a role for OleB, and the last step is directly contradicted by our research in Chapter 6.

OleB

The final gene product of the olefin biosynthesis pathway is OleB, a member of the α/β hydrolase superfamily. The *oleB* gene is always clustered with *oleACD* in olefin-producing bacteria, and it exists as a natural gene fusion with *oleC* in actinobacteria, such as *Arthrobacter and Micrococcus* species (13). Its ubiquitous presence across all 69 bacterial species analyzed strongly suggests a role for OleB in the mechanism for olefin production. Various proposals have been made, such as OleB being an isomerase to convert incorrect stereochemistry of OleA product through keto-enol tautomerization to recover the correct stereochemistry for OleD catalysis, and a structural role in the assembly of a multi-enzyme complex for olefin production. The role of OleB was only recently described as producing the final *cis*-olefin through a novel β-lactone decarboxylase function (20).

The thiolase superfamily of condensing enzymes

Thiolase enzymes are widespread across bacteria, archaea, and eukarya, where they catalyze carbon-carbon bond formation in fatty acid, steroid, and secondary metabolite biosynthesis (14). The Claisen and Claisen-like condensation reactions thiolases catalyze proceed through an acyl-cysteine thioester intermediate. Thiolases are mostly dimeric enzymes that condense two thioester substrates in the form of acyl-CoA or acyl-carrier protein (ACP) (Figure 5A). All enzymes of this superfamily share a common fold consisting of a duplicate, stacked βαβαβαββ motif, one near the amino-terminus and one near the carboxyl-terminus (Figure 5B) (14). This fold aligns loops containing important active site residues, such as the cysteine nucleophile as well as histidine, asparagine, serine, or cysteine residues that act as general bases and/or acids. Thiolase enzymes can be divided into three subcategories: β-ketoacyl-acyl-carrier protein synthases (KAS), polyketide synthases (PKS), and other thiolases such as 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA) synthase and biosynthetic thiolase. A general overview of thiolase enzymes can be found in Table 2. Most KAS (e.g. FabH) and PKS (e.g. chalcone synthase) initiate condensation by decarboxylating the second substrate, creating a nucleophilic carbanion that attacks the thioester of the acyl-enzyme intermediate (Figure 5C) (21,22). Thiolases like HMG-CoA synthase and biosynthetic thiolase instead utilize a general base that deprotonates one substrate, priming it for nucleophilic attack. For HMG-CoA synthase, a proton on C_2 of the acyl-enzyme

intermediate is abstracted by a glutamate, forming a carbanion nucleophile that condenses with the second substrate (Figure 5D) (23). On the other hand, biosynthetic thiolase uses a second cysteine residue as a general base to abstract a proton from C_2 of the second substrate, and condensation proceeds as with KAS and PKS (Figure 5C) (24). Another general difference between thiolase enzymes is the number of substrate coordinating channels needed for turnover. All thiolases share a common channel used to allow substrate to enter the active site and to bind the pantetheine moiety of CoA or ACP. HMG-CoA synthase and biosynthetic thiolase condense relatively short hydrophobic substrates and require only this single channel (Table 2). Thiolases that form longer chain or cyclized products require a second specificity pocket or binding channel, such as most KAS and PKS, but the second substrate is small. It has been suggested that OleA and pyrone synthase require three total channels for proper orientation of substrates for condensation (25,26). Work reported in this dissertation shows the putative three-channel architecture of OleA.

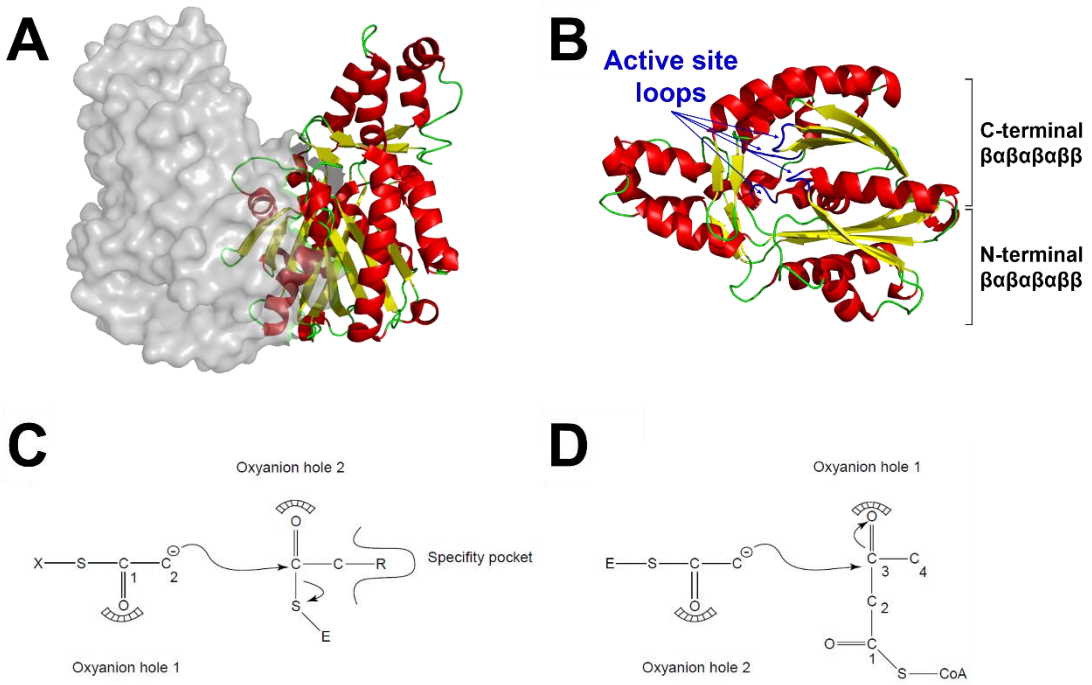


Figure 5: Thiolase enzyme fold and general mechanism of condensation. **A)** Biological assembly dimeric structure of typical thiolase enzyme showing one monomer as a cartoon and the other monomer as a surface model (OleA PDB 3ROW). **B)** The protein fold common among thiolase superfamily members colored by secondary structure (red, α-helix; yellow, β-strand; green, loop; model is from PDB 3ROW) (25). This fold consists of duplicate βαβαβαββ motifs, one near the amino-terminus and one near the carboxyl-terminus of the thiolase. The four loops containing important active site residues are shown in blue. **C)** Thiolase condensation scheme where the second substrate is the nucleophile that attacks the acyl-enzyme intermediate. This mechanistic strategy is typical of KAS, PKS, and biosynthetic thiolase enzymes. **D)** Thiolase condensation scheme where the acyl-enzyme intermediate is the nucleophile that attacks the second substrate. This mechanistic strategy is typical of HMG-CoA synthase and likely pyrone synthase. Panels B and C are from (14).

Table 2: Overview of mechanistic strategies of selected thiolases.

Thiolase functional category	Example enzyme	Number of substrate binding channels	Type of condensation initiation	Nucleophilic substrate
KAS (Types I, II, III)	FabB (Type II)	2	Decarboxylative	Second substrate
	FabH (Type III)	2	Decarboxylative	Second substrate
PKS (Types I, II, III)	Chalcone synthase	2	Decarboxylative	Second substrate
	Stilbene synthase	2	Decarboxylative	Second substrate
	Pyrone Synthase	3[a]	Non-decarboxylative	Acyl-enzyme intermediate
Other	HMG-CoA synthase	1	Non-decarboxylative	Acyl-enzyme intermediate
	Biosynthetic thiolase	1	Non-decarboxylative	Second substrate
	OleA	3[b]	Non-decarboxylative[b]	Second substrate[b]

KAS, β-ketoacyl-acyl-carrier protein synthase; PKS, polyketide synthase; HMG-CoA synthase, 3-hydroxy-3-methylglutaryl-CoA synthase
PDB IDs: FabB, 1FJ8; FabH, 1EBL; chalcone synthase, 1CHW; stilbene synthase, 1U0V; HMG-CoA synthase, 1XPL; biosynthetic thiolase, 1DM3; OleA, 3ROW
[a]Suggested based on homology modelling using OleA as a template (27% sequence identity) (26).
[b]Suggested by publications from this dissertation.

Mechanism and structure of OleA

As the enzyme catalyzes the first committed step of olefin biosynthesis and selects the types of acyl-CoA substrates for condensation, OleA is an important target for enzyme engineering. A detailed mechanistic understanding is needed to understand how substrate is bound, which amino acid residues are important for each chemical step, and how product is released. Knowledge of these details will assist in engineering OleA mutants to select for certain substrates, which when condensed make specific hydrocarbon products.

It was originally proposed that OleA follows a mechanism similar to that of HMG-CoA synthase based on this enzyme being the closest in active site composition (and overall sequence) that has been structurally and mechanistically characterized (7,13). OleA was shown to catalyze a non-decarboxylative Claisen condensation similar to HMG-CoA synthase (15). The proposed mechanism of OleA based on HMG-CoA synthase is illustrated in Figure 6. It was suggested that following transesterification of the first acyl-

substrate to the catalytic cysteine, the covalently bound intermediate becomes activated as a nucleophile for attack of the incoming second acyl-CoA substrate (15). The β-keto acid product would then be released from the cysteine thiolate via a hydrolytic cleavage, likely by an activated water (15).

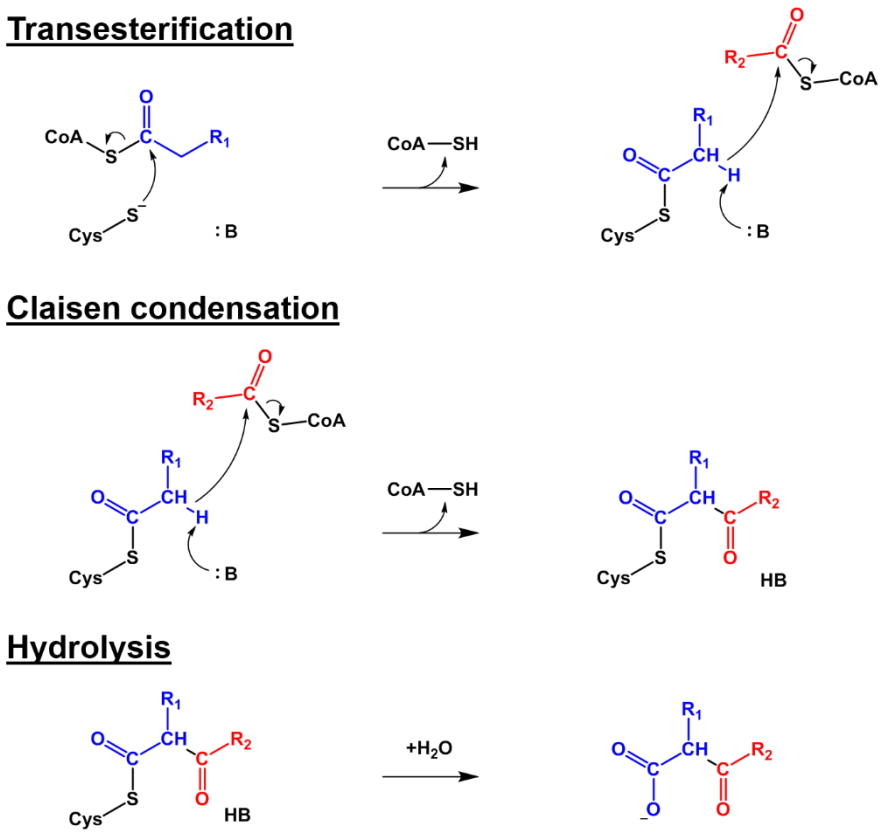


Figure 6: Original mechanism for OleA proposed by (15) and (25). This mechanism proposes that the acylated intermediate (blue) is activated for nucleophilic attack on the incoming second substrate (red) as is the suggested for the mechanism of HMG-CoA synthase (23).

Crystal structures of wild type (WT) and covalently-inhibited OleA from *X. campestris* were solved by Goblirsch and coworkers in 2012 (25). These structures demonstrated a previously undescribed substrate coordinating channel within the active site of the enzyme. The co-crystal structure of WT OleA

with the fungal inhibitor, cerulenin, showed the catalytic cysteine (Cys143) alkylated with cerulenin (25). Surprisingly, the $C_{11:2}$ alkyl chain of this inhibitor was coordinated in a novel binding channel leading from the active site (Figure 7). The depth of this new channel was elucidated by pressurization of OleA-cerulenin crystals with xenon gas. The xenon gas was trapped at the base of this new channel and demonstrated that it extended further than the alkyl chain of cerulenin, consistent with the C_8-C_{16} substrate preference of *X. campestris* OleA (Figure 7B). Other structurally characterized thiolase superfamily enzymes have been shown to require only one or two binding channels in order to coordinate substrate and intermediates for condensation (Table 2). FabH utilizes two channels, one to accommodate the growing alkyl chain and another to allow CoA-activated substrates to enter the active site (Figure 7A) (27). HMG-CoA synthase requires only one channel to bind acetyl-CoA for production of acetoacetyl-CoA (Figure 7A) (28). OleA is unusual in that it must bind two long-chain substrates for condensation. The novel position of cerulenin in the OleA-cerulenin structure allowed for the putative assignment of a three-channel active site centered around Cys143 (Figure 7B) (25).

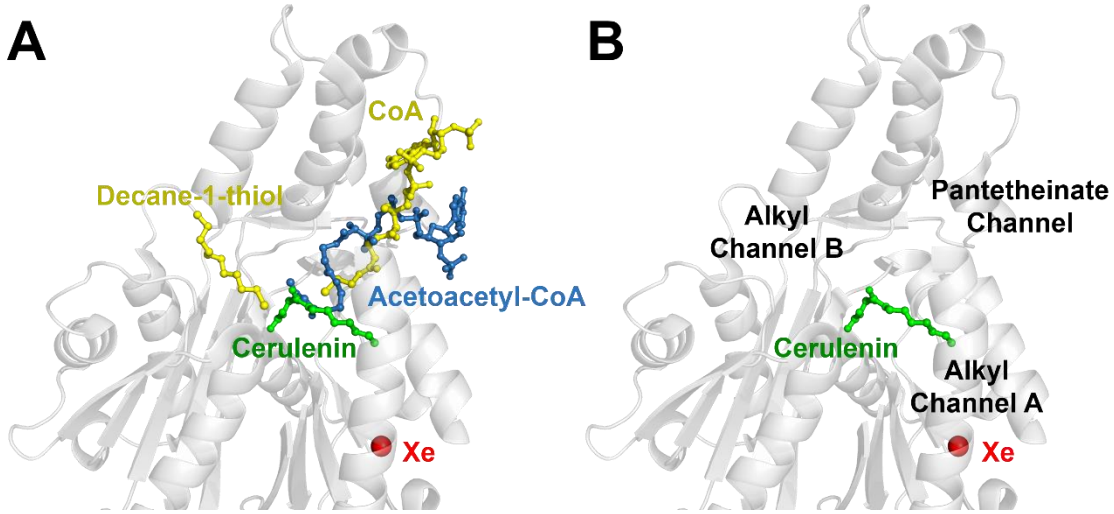


Figure 7: Overlay of thiolase active site substrate coordination. **A)** OleA (gray cartoon) bound with cerulenin (green) overlaid with FabH (bound with decane-1-thiol and CoA, PDB 2QX1) and HMG-CoA synthase (bound with acetoacetyl-CoA, PDB 1TXT) (27,28). Enzyme models for FabH and HMG-CoA synthase have been omitted for clarity. Bound ligands are shown as ball and stick models (yellow and blue, respectively. In OleA, xenon gas (red sphere) is trapped at the base of the channel in which cerulenin is bound. **B)** Putative assignment of OleA binding channels showing cerulenin (green) occupying the novel binding channel. Figure adapted from (25).

Research significance and specific aims

Bacterial production of long-chain olefins represents a viable pathway for renewable and tailored production of specialty chemicals, waxes, and lubricants. A greater understanding of the biochemistry and enzyme structures of this biosynthesis pathway is needed to understand specific hydrocarbons profiles and to facilitate future engineering of products. The overall goals of this dissertation research are to understand the mechanism of OleA (Chapters 2-4), define the true product of OleC and thus infer the role of OleB (Chapter 6), and demonstrate

OleBCD form a multi-enzyme complex to produce hydrophobic olefins (Chapter 7). This dissertation is divided into three specific aims:

1. **Elucidating the mechanism of *X. campestris* OleA**. In order to obtain a more detailed understanding of the mechanism of OleA, X-ray crystallography and mutagenesis of several conserved active site residues are used to assign the roles of these residues during β-keto acid formation.
2. **Characterizing the product of OleC**. Research to define the stereochemical profile of OleC substrates led to the discovery that initial reports of OleC forming *cis*-olefins from β-hydroxy acids by ourselves and others were incorrect. ^{1}HNMR and bioinformatic analyses are used to correctly assign the function of OleC in the olefin biosynthesis pathway, and a preliminary identification of the function of OleB.
3. **Determine the interactions of OleABCD enzymes.** As the substrates, intermediates, and product of the olefin biosynthesis pathway are very hydrophobic, the interactions of OleABCD during olefin production likely play a role in sequestering these molecules from solvent and ensuring these reduced hydrocarbons form the intended olefins. Interactions between these proteins are characterized by chromatography and transmission electron microscopy.

CHAPTER 2

Journal of Biological Chemistry, 2016; 291:26698-26706, 10.1074/jbc.M116.760892.

Substrate Trapping in Crystals of the Thiolase OleA Identifies Three Channels That Enable Long Chain Olefin Biosynthesis

Goblirsch, B. R., Jensen, M. R., Mohamed, F. A., Wackett, L. P. and Wilmot, C. M.

Author contribution

For the work presented in this chapter, Jensen, M.R. optimized and performed all activity assays for OleA and mutant enzymes, modified and edited all figures, and solved the crystal structure for OleA C143S bound with myristic acid and myristoyl-CoA. Jensen, M.R. also helped draft portions of the manuscript and contributed to the final editing of the paper.

Chapter 2 summary

Phylogenetically diverse microbes that produce long chain, olefinic hydrocarbons have received much attention as possible sources of renewable energy biocatalysts. One enzyme that is critical for this process is OleA, a

thiolase superfamily enzyme that condenses two fatty acyl-CoA substrates to produce a β-ketoacid product and initiates the biosynthesis of long chain olefins in bacteria. Thiolases typically utilize a ping-pong mechanism centered on an active site cysteine residue. Reaction with the first substrate produces a covalent cysteine-thioester tethered acyl group that is transferred to the second substrate through formation of a carbon-carbon bond. Although the basics of thiolase chemistry are precedented, the mechanism by which OleA accommodates two substrates with extended carbon chains and a coenzyme moiety—unusual for a thiolase—are unknown. Gaining insights into this process could enable manipulation of the system for large scale olefin production with hydrocarbon chains lengths equivalent to those of fossil fuels. In this study, mutagenesis of the active site cysteine in *Xanthomonas campestris* OleA (Cys143) enabled trapping of two catalytically relevant species in crystals. In the resulting structures, long chain alkyl groups (C_{12} and C_{14}) and phosphopantetheinate define three substrate channels in a T-shaped configuration, explaining how OleA coordinates its two substrates and product. The C143A OleA co-crystal structure possesses a single bound acyl-CoA representing the Michaelis complex with the first substrate, whereas the C143S co-crystal structure contains both acyl-CoA and fatty acid, defining how a second substrate binds to the acyl-enzyme intermediate. An active site glutamate (Gluβ117) is positioned to deprotonate bound acyl-CoA and initiate carbon-carbon bond formation.

Introduction

A diminishing, finite fossil fuel supply has prompted the search for alternative energy platforms and new precursor sources for synthesis of complex, high value commodity chemicals. Production of renewable, energy-rich hydrocarbons and their derivatives is at the epicenter of this effort. Cultivating specialized microbes that produce hydrocarbons represents one attractive option (29). Recently, bacteria across multiple phyla have been shown to be capable of generating long chain, olefinic hydrocarbons ranging in size from C_{23} to C_{33} (7). A total of four olefin (*ole*) biosynthesis genes *oleABCD* are clustered within the bacteria capable of this specialized biosynthesis. Bioinformatic analysis has identified *oleB*, *oleC*, and *oleD* genes as members of the α/β-hydrolase, AMP-dependent ligase, and short chain dehydrogenase superfamilies, respectively (13). The *oleA* gene is a member of the condensing or thiolase enzyme superfamily (10,12).

In vitro studies have demonstrated that OleA, OleD, and OleC enzymes are sufficient for long chain olefin production from acyl-CoA substrate (15). The OleA enzyme initiates biosynthesis by producing a long chain β-ketoacid product that is further processed by OleD to generate a β-hydroxyacid product, with OleC subsequently producing the end point olefin. The exact role of the *oleB* gene product is currently unknown. The OleA, OleD, and OleC enzymes have been individually characterized (15,16,19). Solution studies on *Xanthomonas campestris* OleA revealed the enzyme efficiently catalyzes so called “head-to-

head” condensation between the thioester groups of two long chain fatty acids (C_8-C_{16}) charged in the form of acyl- CoA (15). The mechanism of carbon-carbon bond formation proceeds in a non-decarboxylative fashion between the α-carbon of one substrate to the acyl-carbon of the second (15). The overall reaction generates two free CoA molecules and a single β-ketoacid product that becomes the substrate of OleD (16).

Thiolase superfamily enzymes begin turnover by acylation of an active site cysteine (30). A second substrate then binds, and condensation can proceed. The crystal structure of OleA from *X. campestris* has been determined and consists of the typical homodimeric thiolase protein fold (25). Because OleA catalyzes the condensation of two large substrates with long hydrophobic alkyl moieties, each OleA monomer was predicted to contain three extended substrate channels centered on the active site cysteine (Cys143 in *X. campestris* OleA). The pantetheinate channel, common to all thiolases, would coordinate the phosphopantetheine moiety of each successive acyl-CoA substrate. The two alkyl channels each bind a single alkyl moiety: one associated with the acylated active site cysteine and the other with the second acyl-CoA substrate. In other well studied and structurally characterized thiolase superfamily enzymes, such as β-ketoacyl-acyl carrier protein synthases (27,31,32), polyketide synthases (33), and HMG-CoA synthase (23) at least one of the substrates is small, and so two alkyl channels are not necessary. Of the currently characterized thiolase superfamily enzymes, only OleA and some of the pyrone and pseudopyrone

ketosynthases, such as the recently identified *Pseudomonas* sp. GM30 pseudopyronine synthase, are likely to require three distinct substrate channels (14,26). The previously described crystal structure of OleA in complex with the covalent inhibitor, cerulenin, which has a C_8 hydrocarbon tail, has suggested the position of one of the two alkyl channels (25).

In this study, we have determined the crystal structures of the catalytically impaired C143A and C143S OleA co-crystallized with acyl-CoA substrate. These structures define substrate binding and validate our previously postulated three channel configuration (25). The position of amino acids in the active site relative to substrate coordination suggests a revised mechanism that includes a glutamate donated by the other monomer (Gluβ117 in *X. campestris* OleA): a novel feature of OleA.

Results

Expression and Purification of WT, C143A, and C143S OleA Enzymes

Heterologously expressed C143A and C143S OleA were purified in yields similar to the WT enzyme (15). On average, 1 g of wet cell pellet yielded 750 µg of OleA enzyme. Purity of the sample was confirmed by SDS-PAGE. Electrospray ionization-MS was performed after each purification to ensure the expected mutation had been introduced. The WT, C143A, and C143S OleA had experimental molecular masses of 38,661.8, 38,630.0, and 38,645.8 Da, corresponding well to the predicted masses of 38,661.2, 38,629.2, and 38,645.2 Da, respectively.

Activity of C143A and C143S OleA toward Myristoyl-CoA Substrate

A colorimetric assay using 5,5'-dithio-*bis*-(2-nitrobenzoic acid) (DTNB) was adapted to measure OleA turnover (15,34,35). The assay measures production of free CoA, which can be produced either by proper condensation or futile hydrolysis of the thioester bond of acyl-CoA substrates (Figure 8). *X. campestris* OleA has been previously shown to prefer C_{16} and C_{14} acyl-CoA substrates (15). The C_{14} acyl-CoA substrate, myristoyl-CoA, was selected for DTNB assays to compare WT and mutant OleA turnover. Table 3 summarizes the myristoyl-CoA consumption by WT and OleA enzyme variants. WT OleA enzyme displays robust turnover. Conversely, C143A and C143S OleA display measurable CoA release only after incubation for 24 h.

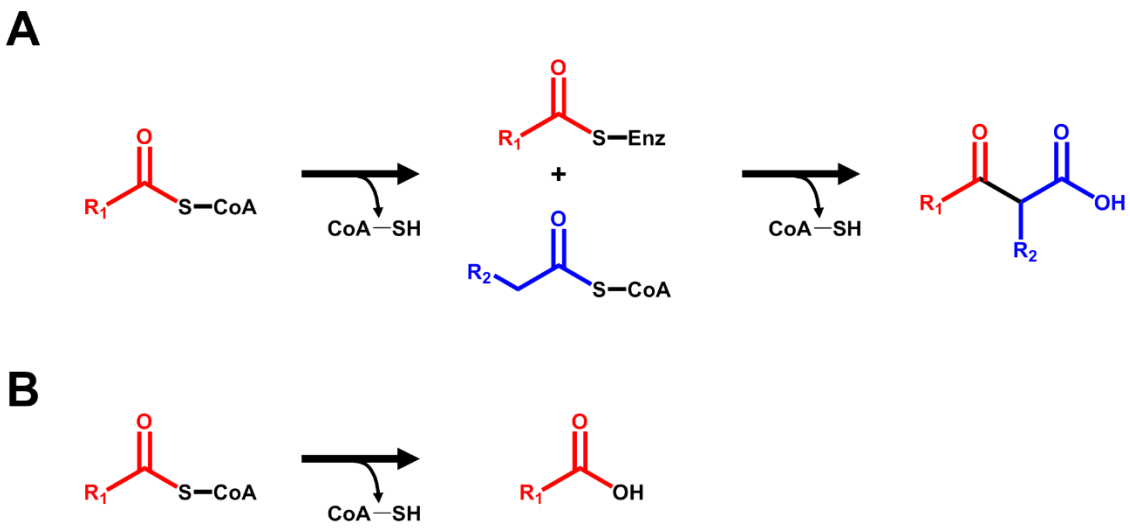


Figure 8: OleA-catalyzed reactions with acyl-CoA substrates. **A)** condensation of acyl-CoA substrates to produce β-ketoacid product. **B)** direct hydrolysis of acyl-CoA substrate.

Table 3: Wild type and mutant OleA consumption of myristoyl-CoA.

Enzyme	Enzyme concentration (μM)	Rxn time (hrs)	CoA product (μM)[a]	Percent yield (%)[b]
WT	0.5	0.25	33 ± 0.5	100
C143A	0.5	0.25	<0.1	<0.1
C143A	0.5	24	0.8 ± 0.4	2.5
C143A	0.5	48	5 ± 0.3	15
C143S	0.5	0.25	0.1 ± 0.07	0.3
C143S	0.5	24	0.4 ± 0.3	1.3
C143S	0.5	48	4 ± 0.08	13

Values shown are the average in triplicate with standard deviation
[a] Free coenzyme A detected as described under "Experimental Procedures"
[b] Starting substrate was 32 μM; 32 μM product is 100% turnover

Overall Structures of OleA Cys143 Variants

The structures of unbound C143A and C143S OleA were determined to 1.85 and 2.36 Å resolution, respectively (Table 4). The gross structure of the homodimer of each variant is nearly superimposable with the WT enzyme structure. The calculated root mean square deviation (rmsd) for all backbone atoms between OleA WT and each variant structure are 0.14 (C143A) and 0.19 Å (C143S). Past attempts at generating WT OleA co-crystal structures with non-covalent ligands (CoA; myristic acid) or trapping in crystallo catalytic intermediates from myristoyl-CoA turnover were unsuccessful, yielding only unbound enzyme. Crystal soaking experiments were also unsuccessful, presumably because of the large ligand and substrate sizes. The minimal activity of C143A and C143S OleA variants with myristoyl-CoA (Table 3) suggested that co-crystallizations were feasible. Cocrystallization of C143A OleA with myristoyl-CoA and C143S OleA with lauroyl-CoA or myristoyl-CoA generated crystals of a

new, bi-pyramidal morphology. Although multiple acyl-CoA lengths were tried in co-crystallization experiments (C_8-C_{16}), only lauroyl-CoA (C_{12} acyl-CoA) and myristoyl-CoA (C_{14} acyl-CoA) produced co-crystals. Furthermore, C143A OleA did not co-crystallize with lauryl-CoA. The structures of C143A OleA-myristoyl-CoA, C143S OleA-myristoyl-CoA, and C143S OleA-lauroyl-CoA were determined to 1.98, 1.97, and 2.17 Å, respectively (Table 4). Overall, the structure of the homodimer of each co-crystal structure is very similar to WT OleA, with the rmsd of all backbone atoms ≤ 0.38 Å (Table 5).

Table 4: C143A and C143S data collection statistics. The data in parentheses are for the highest resolution shell.

	C143A OleA	C143S OleA	C143A OleA–Myristoyl-CoA cocrystal	C143S OleA–Myristoyl-CoA cocrystal	C143A OleA–Lauroyl-CoA cocrystal
Data Collection					
Wavelength (Å)	1.03	1.54	1.03	1.03	1.03
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	84.2 x 85.5 x 103.7	81.9 x 85.5 x 102.0	81.7 x 85.7 x 101.7	81.4 x 85.9 x 101.7	81.4 x 85.5 x 101.6
Resolution (Å)	50.0-1.85 (1.88-1.85)	50.0-2.36 (2.40-2.36)	50.0-1.98 (2.01-1.98)	50.0-1.97 (2.00-1.97)	50.0-2.17 (2.21-2.17)
Measured reflections	463702	211791	415958	390033	263108
Unique reflections	61827	30076	50726	51928	38581
Completeness (%)	98.3 (97.2)	99.7 (96.6)	100 (100)	100 (100)	100 (100)
R_{merge} (%)[b]	7.3 (44.4)	6.3 (33.6)	5.5 (41.7)	7.7 (52.2)	10.9 (57.5)
I/σ_I	23.6 (6.5)	20.5 (3.1)	29.7 (4.7)	23.9 (4.3)	18.2 (3.5)
Multiplicity	7.5 (7.6)	7.0 (5.8)	8.2 (8.2)	7.5 (7.5)	6.8 (6.8)
Wilson B factor (Å^2)	23.0	29.4	28.1	27.2	29.7
Source	APS Sector 23	In house	APS Sector 23	APS Sector 23	APS Sector 23
Refinement					
Resolution (Å)	44.0-1.84 (1.87-1.84)	32.8-2.36 (2.44-2.36)	31.6-1.98 (2.01-1.97)	44.0-1.94 (1.98-1.94)	44.0-2.17 (2.22-2.17)
R_{work}[c]	0.153	0.166	0.157	0.154	0.159
R_{free}[d]	0.200	0.218	0.199	0.196	0.203
Ramachandran statistics (%)					
allowed	100	100	100	100	100
outliers	0	0	0	0	0
RMS deviation					
bond lengths (Å)	0.015	0.003	0.03	0.007	0.009
bond angles (°)	1.546	0.777	1.6940	1.089	1.372
Average B-factor (Å^2)	22.9	37.8	36.0	33.3	34.0
Clashscore[e]	4.23	3.17	4.01	4.59	4.26
Protein Data Bank ID	4KTI	4KTM	4KU2	4KU3	4KU5

[a]Data in parentheses are for the highest resolution shell. [b]$R_{merge} = \Sigma_i |I_{hkl,\,i} - \{I_{hkl}\}/\Sigma_{hkl}\,\Sigma_i\, I_{hkl,\,i}$, where I is the observed intensity and $\{I_{hkl}\}$ is the average intensity of multiple measurements. [c]$R_{work} = \Sigma||F_o|-|F_c|| / \Sigma|F_o|$, where $|F_o|$ is the observed structure factor amplitude and $|F_c|$ is the calculated structure factor amplitude. [d]R_{free} is the R factor based on 5% of the data excluded from refinement. [e]Based on values obtained from MolProbity (http://molprobity.biochem.duke.edu/).

Table 5: Overview of bound channels of reported structures and rmsd[a] from WT OleA.

Crystal	PDB ID	Channel ligands	Channels bound	Rmsd from WT OleA (Å)
C143A	4KTI	None	None	0.14
C143S	4KTM	None	None	0.19
C143A Myristoyl-CoA	4KU2	Myristoyl-CoA	Alkyl channel A, pantetheinate channel	0.32
C143S Myristoyl-CoA	4KU3	Myristoyl-CoA, myristic acid	Alkyl channel A and B, pantetheinate channel	0.38
C143S Lauroyl-CoA	4KU5	Lauroyl-CoA, lauric acid	Alkyl channel A and B, pantetheinate channel	0.32

[a] Root-mean-square-deviation calculated over all backbone atoms between structures by SUPER in PyMol (http://www.pymol.org)

Our previous studies on OleA led to the proposal that three extended substrate binding channels would be required per monomer to complete OleA turnover: one channel dedicated to coordinating the CoA head group and phosphopantetheine arm (pantetheinate channel) and two channels to coordinate the hydrophobic alkyl chains of the acyl-CoA substrates (alkyl channels A and B) (25). The electron density of the C143A OleA myristoyl-CoA structure demonstrated binding of a single myristoyl-CoA molecule (occupancy = 100%) in which the CoA and its phosphopantetheine arm bound as expected, and the C_{14} alkyl chain lay between a β-sheet and two helices to give an L-shaped binding mode (Figure 9A). This channel is undescribed in the thiolase superfamily, but its existence was previously suggested by the structure of WT OleA in complex with the covalent inhibitor, cerulenin, which has a C_8 alkyl tail, and the position of a xenon atom from a xenon-derivative crystal structure of OleA (25). We have termed this channel alkyl channel A. In the C143S OleA-myristoyl-CoA structure, we observed a noncovalently bound myristic acid (occupancy = 100%) within alkyl channel A, whereas a substrate myristoyl-CoA (occupancy = 88%) was positioned within the pantetheinate channel, but with its alkyl chain bound within a channel whose entrance from the active site is directly opposite that of alkyl

channel A (Figure 9B). We term this alkyl channel B. This channel is in the same position as that defined by crystal structures of fatty acid biosynthesis (Fab) β-ketoacyl-acyl carrier protein synthase enzymes; the complex of C112A FabH with lauroyl-CoA (equivalent to C143A *X. campestris* OleA), FabH bound to decane-1-thiol, and FabB bound to cerulenin (21,27,36). Note that in the FabB-cerulenin complex, the C_8 alkyl tail of cerulenin binds in alkyl channel B, whereas in OleA, despite alkyl channelB being present, cerulenin occupies alkyl channel A (25,36). The electron density of C143S OleA-lauroyl-CoA reveals binding of lauric acid (occupancy = 99%) and lauryl-CoA (occupancy = 84%) in positions equivalent to those observed in the C143S OleA-myristoyl-CoA crystal structure (Figure 9C). Overlays between all three co-crystal structures revealed that binding of the CoA portion of all the acyl-CoAs is essentially identical (Figure 10A). Furthermore, the binding of alkyl groups within alkyl channels A and B also occupy comparable positions (Figure 10A). Thus, we can clearly define three extended channels in OleA that together form a T-shaped, three-channel nexus (Figure 10B and Table 5).

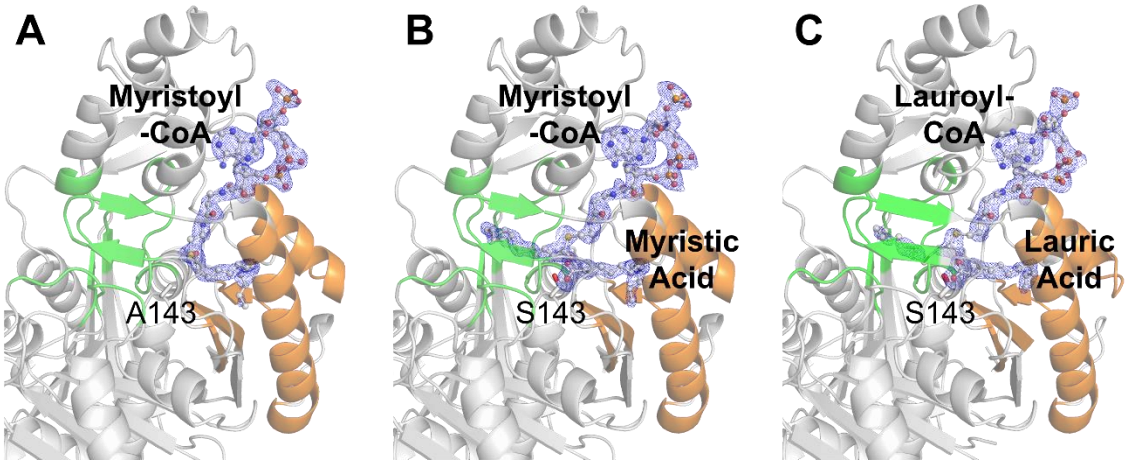


Figure 9: Substrate binding channels in OleA C143 mutants. **A)** C143A OleA bound with myristoyl-CoA (ball and stick). **B)** C143S OleA bound with myristic acid and myristoyl-CoA. **C)** C143S OleA bound with lauric acid and lauroyl-CoA. The single monomer of OleA is shown in *gray cartoon* with alkyl channels A (*orange cartoon*) and B (*green cartoon*) color-coded for clarity. The mutated residue is shown as a stick model (*gray carbon*). Bound substrates are represented as ball and stick models with *blue mesh* illustrating the simulated annealing ligand omit F_o – F_c electron density maps contoured at 2.5 σ.

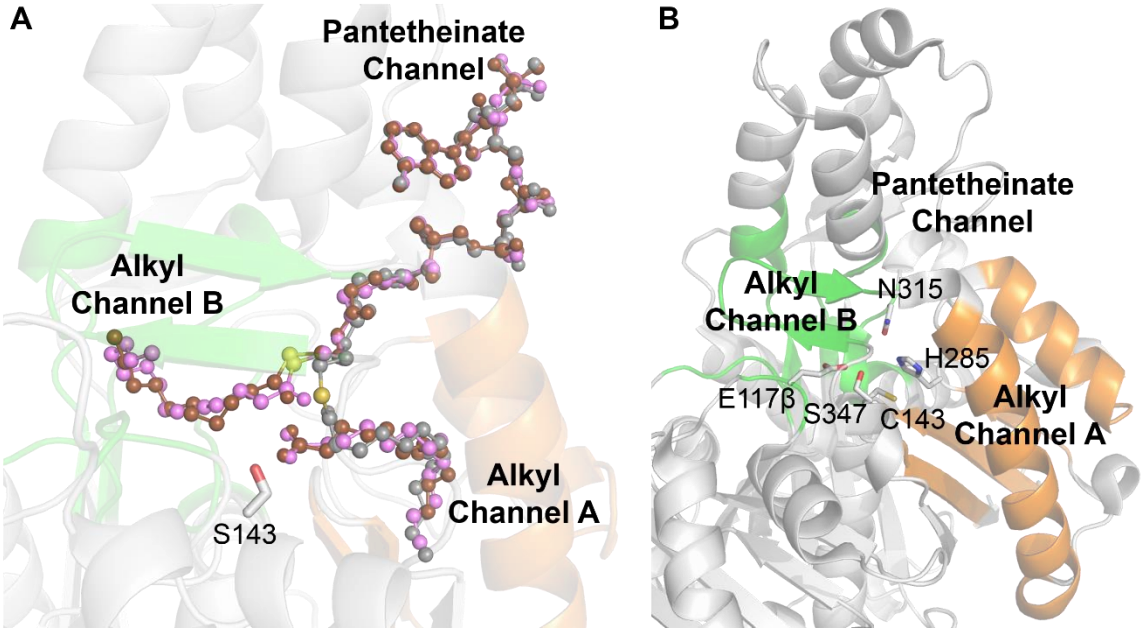


Figure 10: **A)** overlay of C143A and C143S co-crystal structures. Bound ligands (ball and stick) are color-coded by structure: C143A-myristoyl-CoA (*gray*), C143S-myristoyl-CoA (*magenta*), and C143S-lauroyl-CoA (*brown*). Sulfur atoms are colored *yellow* for orientation. **B)** OleA WT monomer highlighting the three-

channel nexus at the active site. Active site residues are shown as sticks (*gray carbon*). Alkyl channels A and B colored as in Figure 9.

The origin of the free fatty acid observed in the C143S OleA complex structures is unclear but likely comes from hydrolysis of a species bound in alkyl channel A. One possibility is hydrolysis of a covalent acyl intermediate generated by ester bond formation between acyl-CoA and deprotonated Ser143. An ester-containing covalent acyl intermediate has previously been observed in an equivalent Cys to Ser variant of the thiolase *Escherichia coli* β-ketoacyl-ACP synthase I (37). Such covalent acyl intermediates have enhanced stability (ester *versus* thioester) and are often observable. However, in the case of C143S OleA, efforts to observe such a species by electrospray ionization-MS or through co-crystallization have failed. The second possibility is hydrolysis of acyl-CoA substrate non-covalently bound in the active site. Regardless of the species hydrolyzed, the end result is a fatty acid tightly bound in alkyl channel A that does not dissociate before a second acyl-CoA binds. Of note is that only one monomer active site of the OleA dimer in the crystallographic asymmetric unit contains bound fatty acid and/or acyl-CoA in all three OleA variant co-crystal structures, with the other monomer active site showing no evidence of significant binding. Because these are co-crystal structures, this means that dimers in which only one monomer has bound to substrate were selected out by the crystallization process. In contrast, the previously reported WT OleA-cerulenin co-crystals crystallized in a different space group with only a monomer in the asymmetric

unit, and thus inhibitor was equally present in both monomers of the dimer generated by crystallographic symmetry (25).

Interactions of Acyl-CoA within the OleA Pantetheinate Channel

The position of the CoA moiety in all acyl-CoA cocrystal structures is conserved (Figures Figure 9 and Figure 10A). Ionic and polar interactions occurring between myristoyl-CoA and C143A OleA are shown as representative of all three structures (Figure 11A). The amine group of the myristoyl-CoA adenine ring forms hydrogen bonds with the hydroxyl group of Thr250 and the main chain carbonyl oxygen of Arg195, whereas the 3’ phosphate group forms a hydrogen bond with the hydroxyl group of Thr60. The guanidinium group of Arg195 also forms a cation-π stacking interaction with the adenine ring of the acyl-CoA. In OleA, the phosphopantethine group of the acyl-CoA forms no discernable hydrogen bond or ionic contacts with the remainder of the pantetheinate channel leading to the active site.

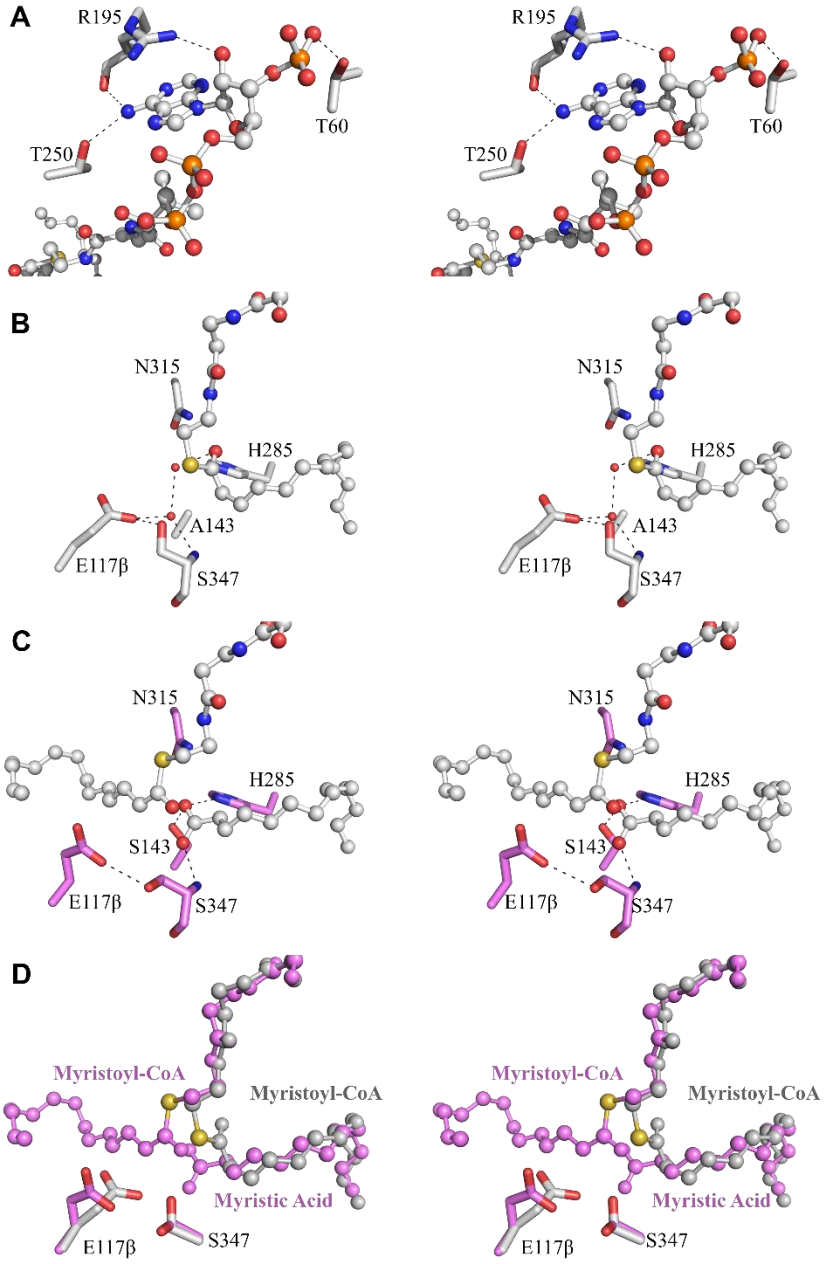


Figure 11: Stereoviews (cross-eyed) of the C143A and C143S myristoyl-CoA co-crystal active sites. **A)** acyl-CoA bound in the OleA pantetheinate channel. Pantetheinate channel residues forming hydrogen bonds with the CoA moiety are shown as sticks (*gray carbon*). The guanidium group of residue Arg195 additionally pi-stacks with the adenine ring of coenzyme A. **B)** C143A bound with myristoyl-CoA (ball and stick). Active site residues are shown as sticks (*gray carbon*). **C)** C143S bound with myristic acid and myristoyl-CoA (ball and stick) occupying alkyl channelsAand B, respectively. Active site residues are shown as sticks (*pink carbon*). **D)** Overlay of C143A and C143S co-crystal structures. Color

coding for active site residues are conserved as in B and C. Bound myristoyl-CoA (*gray*, C143A co-crystal; *pink*, C143S co-crystal) and myristic acid (*pink*, C143S co-crystal) are explicitly labeled. Note significant movement of the side chains of Gluβ117 and Ser347 occurs to accommodate the alkyl chain in channel B. In all panels, hydrogen bond contacts are represented by *dashed lines*, and the ordered waters are shown as *red spheres*.

Active Site Perturbations Induced by Acyl-CoA and Fatty Acid Binding

The active site of OleA features a catalytic cysteine residue (Cys143) lying at the terminus of the pantetheinate channel (Figure 10B). Residues are positioned immediately above Cys143 to help stabilize tetrahedral, oxyanionic intermediates formed during OleA turnover. Pockets formed by the side chains of His^{285} and Asn^{315} and the main chain amides of Cys143 and Ser347 stabilize these intermediates. A glutamic acid (Gluβ117) is also positioned in the active site but unusually originates from the neighboring monomer. Mutation of the active site Cys143 of OleA causes no perturbation of active site residues in the C143A and C143S OleA variant structures. When myristoyl-CoA is bound to C143A OleA, minimal perturbation is observed (Figure 11B). Gluβ117 is positioned in the active site, and hydrogen bonds with the hydroxyl of Ser347 and an active site water molecule. Interestingly, the thioester moiety of myristoyl-CoA forms no hydrogen bonds with OleA active site residues and is positioned 5.5Å away from the Cβ of Ala143. In the C143S-myristoyl-CoA crystal structure, alkyl channel A is occupied by myristic acid, and a myristoyl-CoA molecule is stacked above with its alkyl chain lying in alkyl channel B (FiguresFigure 9B and Figure 11C). In this structure, active site waters are completely displaced. The myristic acid carboxylate forms an extensive hydrogen bonding network with the side

chains of Ser143 and His285 and the main chain amide of Ser347. Gluβ117 remains hydrogen bonded to Ser347, but the presence of the alkyl chain from myristoyl-CoA forces an outward rotation of the carboxylate (Figure 11D).

Alkyl Chain Binding Environments in Channels A and B

Alkyl channel A is formed by a pair of α-helices composed of residues 252–272 and 289–301 flanked by a region of β-sheet consisting of residues 280–284, 339–346, and 349–355. The C143A-myristoyl-CoA and C143S-myristoyl-CoA structures have long chain alkyl groups (C_{14}) bound in alkyl channel A. These are highly kinked with their terminal carbon positioned near Cys351, Met353, and Phe295 (Figure 12A). Binding of the hydrocarbon chain in alkyl channel A results in little perturbation with the exception of two amino acid residues. His291, positioned at the ceiling of the channel, flips out of the pocket and forms a hydrogen bond interaction with Ser288 (Figure 12B). At the base of the channel, the Ile345 residue also adopts an alternative rotamer, and the peptide backbone moves deeper into the base of the channel (Figure 12B). The path of the alkyl chain is being driven by the side chain positions of Ile258, Ile284, Val287, and Phe295, which curls the alkyl moiety around the Ile345 side chain.

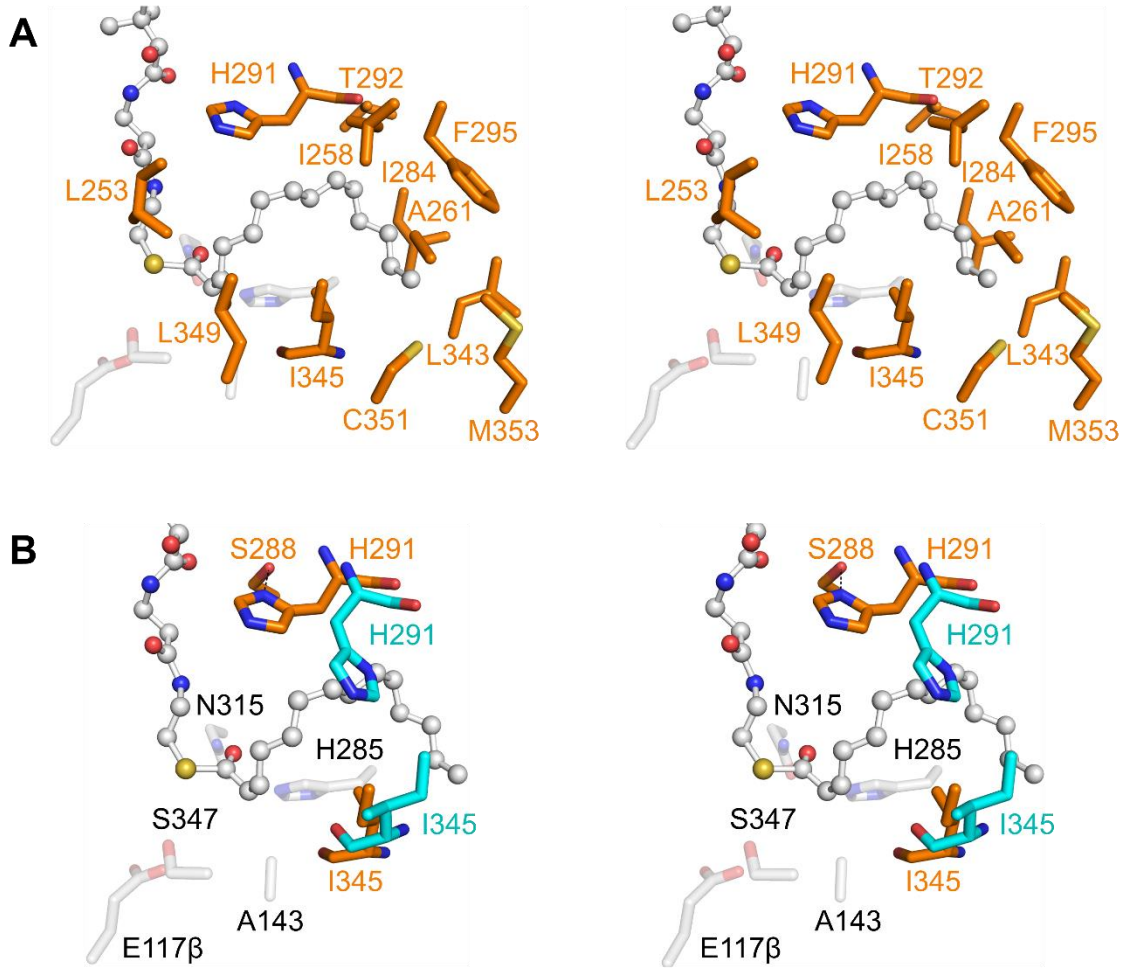


Figure 12: Stereoviews (cross-eyed) of OleA alkyl channel A. **A)** C143A bound with myristoyl-CoA (ball and stick). Residue side chains forming the channel are indicated by sticks (*orange carbon* and *label*). **B)** Overlay of unbound C143A OleA (*blue carbon*) with C143A co-crystallized with myristoyl-CoA (*orange carbon*). Residues His291 and Ile345 undergo significant conformational changes to bind myristoyl-CoA. Active site residues are shown as faded sticks (*gray carbon*). Hydrogen bond contacts are represented by *dashed lines*.

Residues originating from multiple structural elements form alkyl channel B. Two α-helix termini formed by residues 170– 177 and 201–207 and three loops consisting of residues 110–114, 239–249, and 315–318 compose the majority of the channel (Figure 10B). A loop from the other monomer of the dimer containing Gluβ117 (residues 116–118) completes alkyl channel B. To fit within

the channel, the C_7-C_{14} carbons of the myristoyl-CoA alkyl group curve back toward the OleA active site with the final alkyl carbon positioned near Met246 and Gluβ117 (Figure 13). The base and rear of alkyl channel B also possesses residues with charged/polar side chains that may also encourage the curving of the alkyl chain back upon itself.

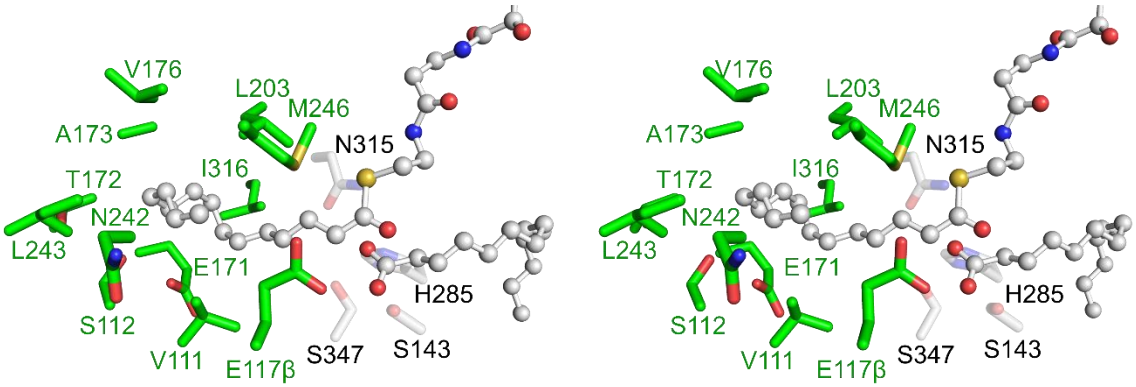


Figure 13: Stereoview (cross-eyed) of OleA alkyl channel B. C143S OleA bound with myristoyl-CoA (ball and stick) within alkyl channel B. Residue side chains forming the channel are indicated by sticks (*green carbon*). Active site residues are shown as faded sticks (*gray carbon*).

Discussion

Intermediate binding steps observed during OleA turnover have previously proved elusive. Whereas the general mechanistic outlines of carbon-carbon bond formation had been established, the events surrounding the condensation step were not well understood (12,15). The non-decarboxylative Claisen mechanism of condensation catalyzed by OleA requires two thioester-acyl groups to be coordinated within the active site simultaneously prior to condensation. A three-channel nexus at the active site was proposed to enable the condensation chemistry to commence (25). The structures presented here validate this model,

and allow for the proposal of events before and after condensation. The C143A and C143S OleA variants provided the opportunity to trap acyl-CoA substrate within OleA and provide insight into the workings of these essential channels (Figure 10 and Table 5)

As in all thiolases, a single pantetheinate channel leads to the active site of OleA. Although the CoA head group makes multiple ionic and hydrogen bond interactions at the surface of OleA, the phosphopantetheine arm of the CoA makes no charged/polar interactions through the remainder of the channel leading to the OleA active site (Figure 11A). This is consistent with the need for uncharged CoASH to have a high dissociation constant enabling efficient release and allowing entry of the second acyl-CoA substrate. Analogously, FabH makes few contacts with the CoA moiety of bound lauroyl-CoA, where release of CoASH is required for the binding of the second substrate, malonyl-ACP (21,38).

Alkyl channels A and B are perpendicular to the pantetheinate channel and lead away from the active site to create a T-shaped configuration of channels (Figure 10). OleA represents the first description of alkyl channel A, and its presence in other thiolases with two bulky substrates is likely. In particular, some of the ketosynthases involved in bacterial pyrone biosynthesis create an α-pyrone ring with large orthogonal substituents, including the antibiotics myxopyronin, corallopyronin, and pseudopyronin (26,39,40). The orthogonal nature of the substituents matches well with the orthogonal alkyl channels observed in OleA, as was noted in the recent crystal structure of *Myxococcus*

fulvus Mx f50 MxnB, the stand-alone ketosynthase that completes myxopyronin biosynthesis (41). The alkyl chains in both OleA alkyl channels A and B are not linear but curled, explaining how different lengths of alkyl chain can be accommodated in the active site and why *X. campestris* OleA is promiscuous (13). The residues lining these channels are not conserved in OleAs, as would be expected given the range of substrate specificity across the enzyme family. A final note is that in both the WT OleA-cerulenin (25) and C143A-myrisotyl-CoA structures, alkyl channel A is occupied first (Figure 9A), consistent with this binding the first substrate and thus the covalent acyl-enzyme intermediate prior to condensation. Interestingly, Sucipto *et al.* (41) suggested that in MxnB the first substrate would bind in alkyl channel B based on the steric constraints required for α-pyrone ring formation.

The active site of OleA experiences few changes upon acyl-CoA or acyl-CoA and fatty acid binding (Figure 11). The bound myristoyl-CoA within C143A OleA rests above the active site residues (Figure 11B). Only solvent molecules remain between the thioester group of myristoyl-CoA and Ala143, suggesting that in WT OleA, descent of myristoyl-CoA to Cys143 could readily occur to enable transeserification. The presence of Ser143 in the C143S OleA introduces an additional hydrogen bond donor into the active site, which participates in hydrogen bonds to the non-physiological myristic acid head group created by offpathway hydrolysis (Figure 11C). The binding of the second myristoyl-CoA substrate requires occupation of alkyl channel B and positions the second

substrate above the carboxylate of myristic acid. Myristoyl-CoA binding also induces movement of residue Gluβ117 and its hydrogen bond partner Ser347 (Figure 11D). The observed stacking of myrisotyl-CoA above myristic acid in OleA mimics the complex between the covalent acyl-OleA intermediate and the second acyl-CoA substrate. Like OleA, HMG-CoA synthase also proceeds through a non-decarboxylative mechanism for carbon-carbon bond formation and contains the same spatially conserved residues in the active site (15,42). In HMG-CoA synthase, the proton abstraction that initiates carbon-carbon bond formation is catalyzed by an active site glutamate, which is 3.0 Å from the C_2 of the acylated Cys but 5.7 Å from the C_3 of the second substrate acetoacetyl-CoA (23). The carboxylate of Gluβ117 in OleA is spatially equivalent to the HMG-CoA glutamate, but because the alkyl chain of the first substrate occupies channel A, the relevant carbon belonging to the acylated Cys143 is now more distant (7.1 Å based on the myristic acid position in the C143S OleA variant complex), and the angle between the carboxylate and the hydrogen makes proton abstraction from this group improbable. However, the carboxylate of Gluβ117 is 4.0 Å away from the α-carbon of the acyl group in alkyl channel B, suggesting that OleA abstracts the proton from the second, non-covalently bound substrate. The interaction with Ser347 likely plays a role in activating Gluβ117, promoting abstraction of the acidic proton from the second substrate. The created carbanion would then perform a nucleophilic attack on the acyl-enzyme intermediate to generate the carbon-carbon bond (Figure 11C). In corroboration of this idea, studies on

Kineococcus radiotolerans OleA (39% identity with *X. campestris* OleA) have identified a β-ketoacyl-CoA intermediate preceding β-ketoacid formation, which could only occur if the second substrate is activated for nucleophilic attack on the acylated enzyme (16). The combination of the structural data presented here and previous *in vitro* studies leads to the proposal of an updated mechanistic model for OleA turnover (Figure 14).

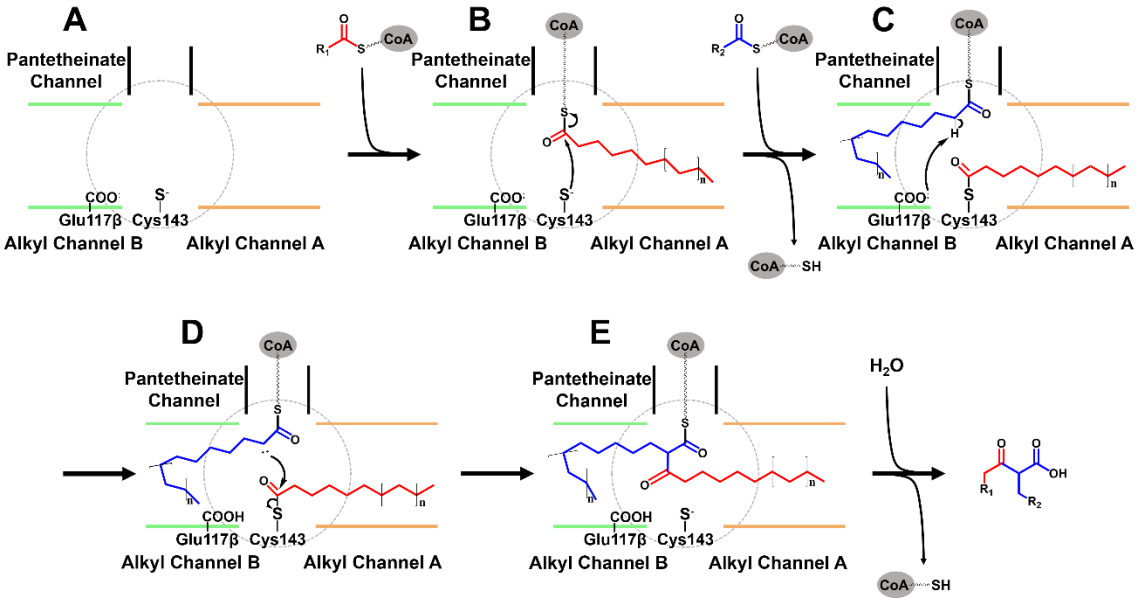


Figure 14: Substrate coordination and proposed reaction scheme of OleA. The active site of a single monomer is depicted with three substrate channels labeled. Alkyl channel A (*orange*) and alkyl channel B (*green*) are color-coded to be consistent with all previous figures. Coenzyme A is represented by a *gray oval* connected to the phosphopantetheine arm depicted as a *wavy line* terminating in the reactive thiol group. **A)** Resting state of OleA. **B)** Binding of the first acyl-CoA substrate within alkyl channel A prior to transesterfication. The variable length of the alkyl group is signified by *n* repeating units. **C)** Binding of the second acyl-CoA substrate within alkyl channel B prior to C–C bond formation. Note a base, putatively assigned as Gluβ117, required to preempt C–C bond formation by proton abstraction. **D)** Nucleophilic attack by the α-carbanion of the second acyl-CoA on the thioester of the enzyme-acyl intermediate to form a C–C bond. **E)** Bound β-ketoacyl-CoA prior to hydrolysis. The release of free coenzyme A and β-ketoacid product complete OleA turnover back to the resting state.

Dissociation of the β-ketoacid product is non-trivial because it is so large. It seems likely that the product exits from alkyl channel B, because the β-hairpin (residues 239–249; Figure 10) that shields the alkyl chain from solvent shows enhanced mobility in the absence of bound ligand (25). The OleA homodimer possesses a pseudo-2-fold axis that places both β-hairpin loops from the monomers in a deep surface groove that contains elements of α-helices in other thiolase superfamily structures (Figure 15). OleA produces a hydrophobic β-ketoacid product that is labile and unlikely to be released into the cytosol (15). Therefore, the open interface of the OleA homodimer raises the intriguing possibility that the subsequent olefin biosynthesis enzyme, OleD, could directly dock with OleA to facilitate product release and acquire the β-ketoacid.

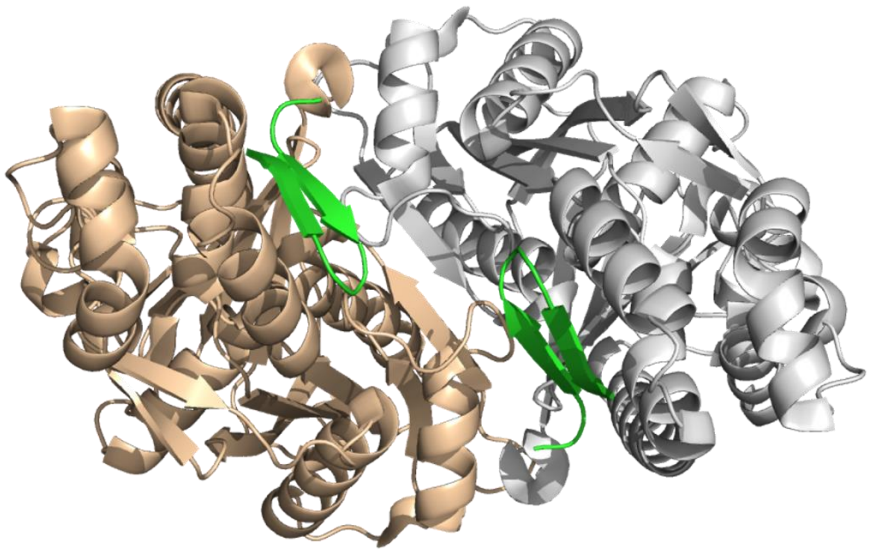


Figure 15: Top-down view of the OleA dimer shown as schematic with each monomer colored coded as *tan* or *gray*. The loop (residues 238–248) that lies over alkyl channel B is explicitly shown as *green schematic*. The image orientation is 90° rotation about an axis horizontally in the plane of the paper from the orientation in Figure 9.

Experimental Procedures

Preparation of OleA Site Mutants C143A and C143S OleA

The cloning of the synthetic *oleA* gene from *X. campestris* spv. campestris str. ATCC 33913 (NP_635607.1) into pET28b (Novagen, Madison, WI) was described previously (15). Site specific primers (Integrated DNA Technologies) were designed to mutate the Cys143 amino acid in *X. campestris* OleA to a corresponding C143A or C143S. Phusion polymerase (New England Biolabs) was used to generate the mutants via PCR (43). The parental template was removed by digestion with DpnI. Mutagenesis to either C143A or C143S was confirmed by DNA sequencing (ACGT Inc.).

Purification of C143A and C143S OleA

Cultivations of WT OleA and OleA mutants were conducted in 2-liter flasks containing 1 liter of LB medium with 50 μg/ml kanamycin. The BL21 *E. coli* cells (Invitrogen) were induced at an *A*600 of 0.6–0.8 with 0.1 mM isopropyl-β-D-thiolgalactopyranoside. After 4 h, the cells were harvested by centrifugation for 15 min at 4000 × *g*.

The cells were resuspended in 500 mM NaCl, 20 mM sodium phosphate buffer, pH 7.4, with EDTA-free protease inhibitor tablets (Roche Applied Science). The cells were lysed via a chilled French pressure cell at 20,000 p.s.i. and centrifuged at 27,000 × *g* for 70 min to obtain the soluble fraction. The soluble fraction was then run through a 0.45-μm syringe filter prior to loading onto a ÄKTA FPLC (General Electric) equipped with a 5-ml Ni(II)-loaded His-Trap

column equilibrated with 500 mM NaCl, 30 mM imidazole, 20 mM sodium phosphate buffer, pH 7.4. WT OleA and C143A/S point mutants all eluted between 130 and 160 mM imidazole, as monitored by absorbance at 280 nm. Enzyme purity was judged by SDS-PAGE and SimplyBlue Safestain (Invitrogen). C143A and C143S OleA variants were further confirmed by mass spectrometry using an Ion-Spray electrospray source coupled to a QSTAR XL (AB Scienx) quadrupole TOF mass spectrometer. All OleA samples were concentrated to between 225 and 300 µM for crystallization trials.

Detection of OleA Turnover of Myristoyl-CoA Substrate

Consumption of myristoyl-CoA (Sigma) substrate during OleA turnover was monitored through quantification of the free thiol group of CoASH product. In the assay, CoASH reacts with DTNB, releasing chromophoric 2-nitro-5-thiobenzoic acid anion, which is measured spectrophotometrically at 412 nm (ε_{412} = 13,600 M^{-1} cm^{-1}) (15,34,35). OleA enzyme (20 µg) was reacted with 65 µM myristoyl-CoA in 490 µL containing 200 mM sodium chloride, 20 mM Tris HCl, pH7.4, with 4% (v/v) ethanol, and incubated at room temperature for times ranging from 5 min to 48 h. Each reaction was quenched rapidly by the addition of 4.1 M guanidine HCl at pH 7.4, followed by DTNB addition for 10 min, and then quantification via UV-visible spectrophotometry. Enzyme-dependent consumption of myristoyl-CoA was quantified using the Beer-Lambert Law, controlling individually for background enzyme and substrate (non-enzymatic hydrolysis) contributions to absorbance.

Crystallization of C143A and C143S OleA

Crystallization of WT OleA has been reported previously (25). The C143A and C143S OleA crystals grew from similar conditions. The crystals were grown by hanging drop vapor diffusion at 20°C. Drops consisted of 1 µL of protein and 1 µL of mother liquor solution (15–20% (w/v) PEG 8000, 70–100 mM potassium phosphate dibasic, and 100 mM sodium citrate, pH 4.2). The crystals were cryoprotected in mother liquor containing 20% PEG 400. Unbound C143A and C143S OleA crystals appeared in 1–3 days and had a rod-like morphology identical to previously characterized WT OleA crystals grown under the same conditions.

C143A and C143S co-crystals were generated by reacting the variant enzyme with 1 mM acyl-CoA substrate (Sigma) for 1 h at 4 °C. Excess acyl-CoA caused phase separation in the crystallization drops, so excess acyl-CoA was removed by a dialysis step prior to setting up the crystallization trays. Co-crystals appeared in 1–4 days with a bi-pyramidal morphology distinct from the unbound OleA crystals.

X-ray Data Collection, Processing, and Refinement

Using 1.03 Å wavelength X-rays, diffraction data sets were collected at National Institute of General Medical Sciences and National Cancer Institute Collaborative Access Team Beamlines 23-ID-D and 23-ID-B of the Advanced Photon Source (APS) (Argonne National Laboratory, Argonne, IL) at 100 K using a MARmosaic 300 CCD. Beam size was adjusted to match the crystal size and

orientation and attenuated by 50–150-fold. The data set using X-rays with a wavelength of 1.54 Å was collected on a copper rotating anode X-ray generator (Rigaku) with a Saturn 944 CCD detector at 100 K at the Kahlert Structural Biology Lab, University of Minnesota. The data collection statistics are listed in Table 4.

All C143A and C143S OleA crystals belonged to space group $P2_12_12_1$ and were isomorphous to the previously reported structure of unbound WT OleA (Protein Data Bank entry 3ROW) (25). The asymmetric unit of the unbound WT OleA consists of a homodimer, and this was used as the initial phasing model in difference Fourier synthesis to solve all mutant structures with PHENIX 1.8.2–1309 (44). The designated *R*free reflections were carried over from theWTOleA data set. For all structures, cycles of model building with COOT (45) and restrained refinement using PHENIX with TLS were performed until all interpretable regions of the $2F_o - F_c$ and $F_o - F_c$ were explained. All $F > 0\sigma F$ were used in refinement. Refinement statistics were validated using MolProbity (46). All figures illustrating enzyme structures were prepared using PyMOL (47).

CHAPTER 3

This chapter (with supplemental material reintegrated) is reprinted with permission from Portland Press *Biochemical Journal*.

Biochemical Journal, 2017; 10.1042/BCJ20170642.

OleA Glu117 is key to condensation of two fatty-acyl coenzyme A substrates in long-chain olefin biosynthesis

Jensen, M.R., Goblirsch, B.R., Christenson, J.K., Esler, M.A., Mohamed, F.A., Wackett, L.P., Wilmot, C.M.

Chapter 3 summary

In the interest of decreasing dependence on fossil fuels, microbial hydrocarbon biosynthesis pathways are being studied for renewable, tailored production of specialty chemicals and biofuels. One candidate is long-chain olefin biosynthesis, a widespread bacterial pathway that produces waxy hydrocarbons. Found in three- and four-gene clusters, *oleABCD* encode the enzymes necessary to produce *cis*-olefins that differ by alkyl chain length, degree of unsaturation, and alkyl chain branching. The first enzyme in the pathway, OleA, catalyzes the Claisen condensation of two fatty acyl-coenzyme A molecules to form a β-keto acid. In this report, the mechanistic role of *Xanthomonas campestris* OleA Glu117 is investigated through mutant enzymes. Crystal structures were determined for each mutant as well as their complex with the

inhibitor cerulenin. Complemented by substrate modelling, these structures suggest that Glu117 aids in substrate positioning for productive carbon-carbon bond formation. Analysis of acyl-coenzyme A substrate hydrolysis shows diminished activity in all mutants. When the active site lacks an acidic residue in the 117 position, OleA cannot form condensed product, demonstrating Glu117 has a critical role upstream of the essential condensation reaction. Profiling of pH dependence shows that the apparent pK_a for Glu117 is impacted by mutagenesis. Taken together, we propose that Glu117 is the general base needed to prime condensation via deprotonation of the second, non-covalently bound substrate during turnover. This is the first example of a member of the thiolase superfamily of condensing enzymes to contain an active site base originating from the second monomer of the dimer.

Introduction

The transition from reliance on fossil fuels to developing renewable hydrocarbons for biofuels and commodity chemicals has been a major focus in recent biotechnology research (4,48,49). Microbial production of specialty hydrocarbons represents a viable option for this transition (48). Much research has been conducted to identify candidate organisms for heterologous expression and synthesis of these compounds; however, a great deal is yet unknown about how and why the native organisms produce these hydrocarbons.

One such platform of microbial hydrocarbon production is long-chain (C_{23}-C_{33}) olefin biosynthesis. These waxy hydrocarbons were first characterized structurally in 1969, and they have been described as arising from a head-to-head condensation of two fatty acids, leading to an odd-numbered alkene with a double bond at the site of condensation (8,13). The genes and enzymes responsible for olefin biosynthesis have been elucidated across deeply rooted branches of the prokaryotic tree of life (9,13). The enzymes encoded by four clustered genes in these bacteria – *oleABCD* – comprise the Ole pathway. The first enzyme in the pathway, OleA (thiolase superfamily), catalyzes the Claisen condensation of two fatty acyl-coenzyme A (CoA) molecules to form a β-keto acid (15,25). OleD (short-chain dehydrogenase/reductase superfamily) reduces the β-keto acid to a β-hydroxy acid in an NADPH-dependent reaction (16). Recently, it was demonstrated that OleC (Acyl-CoA synthetase-Nonribosomal peptide synthetase adenylation domain-Luciferase superfamily) functions as a β-lactone synthetase, using ATP hydrolysis to form a β-lactone (50). It is hypothesized that the final decarboxylation to form the *cis*-olefin product is catalyzed by OleB (α/β-hydrolase superfamily) (50). OleBCD were proposed to form a multi-enzyme assembly that may function to efficiently move the hydrophobic pathway intermediates between enzyme active sites, preserve the stereochemistry, and sequester the highly reactive β-lactone from the cell (51). As well as initiating the pathway, OleA has been proposed to act as a shuttle

between fatty acid metabolism and the OleBCD assembly feeding its β-keto acid product to OleD (51).

Most thiolase superfamily enzymes exist as homodimers with a subset forming homotetrameric (dimer of dimers) structures (14,52). Thiolases can be classified into at least three distinct functional subcategories: β-ketoacyl-acyl-carrier protein synthases (KAS), polyketide synthases (PKS), and biosynthetic thiolases/3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) synthases (14). All superfamily enzyme mechanisms proceed through a thioester acyl-enzyme intermediate with a conserved active site cysteine (acetyl transfer/transesterification) (14). After binding a second substrate, a thioester-dependent Claisen condensation occurs through nucleophilic attack by a carbanion at the C2 position of the nucleophilic substrate. Thiolases such as FabH (KAS type III) generate a carbanion on C2 of the acyl carrier protein-bound second substrate by promoting decarboxylation (21). On the other hand, biosynthetic thiolase and HMG-CoA synthase generate the carbanion nucleophile in a non-decarboxylative fashion by deprotonation of one substrate by a catalytic base, a cysteine or glutamate, respectively (24,53). The final step involves product release (KAS, PKS, biosynthetic thiolase) or hydrolysis of a thioester-linked product (HMG-CoA synthase). A summary of thiolase superfamily enzyme mechanisms can be found in Figure 16. It has been shown that OleA condenses two long-chain fatty acyl-CoA substrates using a non-decarboxylative mechanism (15). Crystal structures of wild type (WT) OleA and

Cys143 mutants co-crystallized with substrate revealed that OleA has a novel three-channel active site architecture (

Figure **17**) (54). The pantetheinate channel coordinates the phosphopantetheine moiety of the acyl-CoA substrates while two alkyl channels coordinate the acyl-enzyme intermediate and second acyl-CoA substrate in alkyl channels A and B, respectively (54). Other structurally characterized thiolase superfamily enzymes, such as FabH, PKS type II, and HMG-CoA synthase condense smaller substrates and only require two (FabH and PKS type II) or one (HMG-CoA synthase) substrate binding channels, respectively (21,33,53,55). The only thiolase superfamily enzymes known to require three distinct channels are OleA and pyrone and pseudopyrone ketosynthases (41,54). It has been proposed that these enzymes also utilize an active site general base, a glutamate, that comes from the second subunit of the dimer to initiate Claisen condensation, but this has never been demonstrated (26,54).

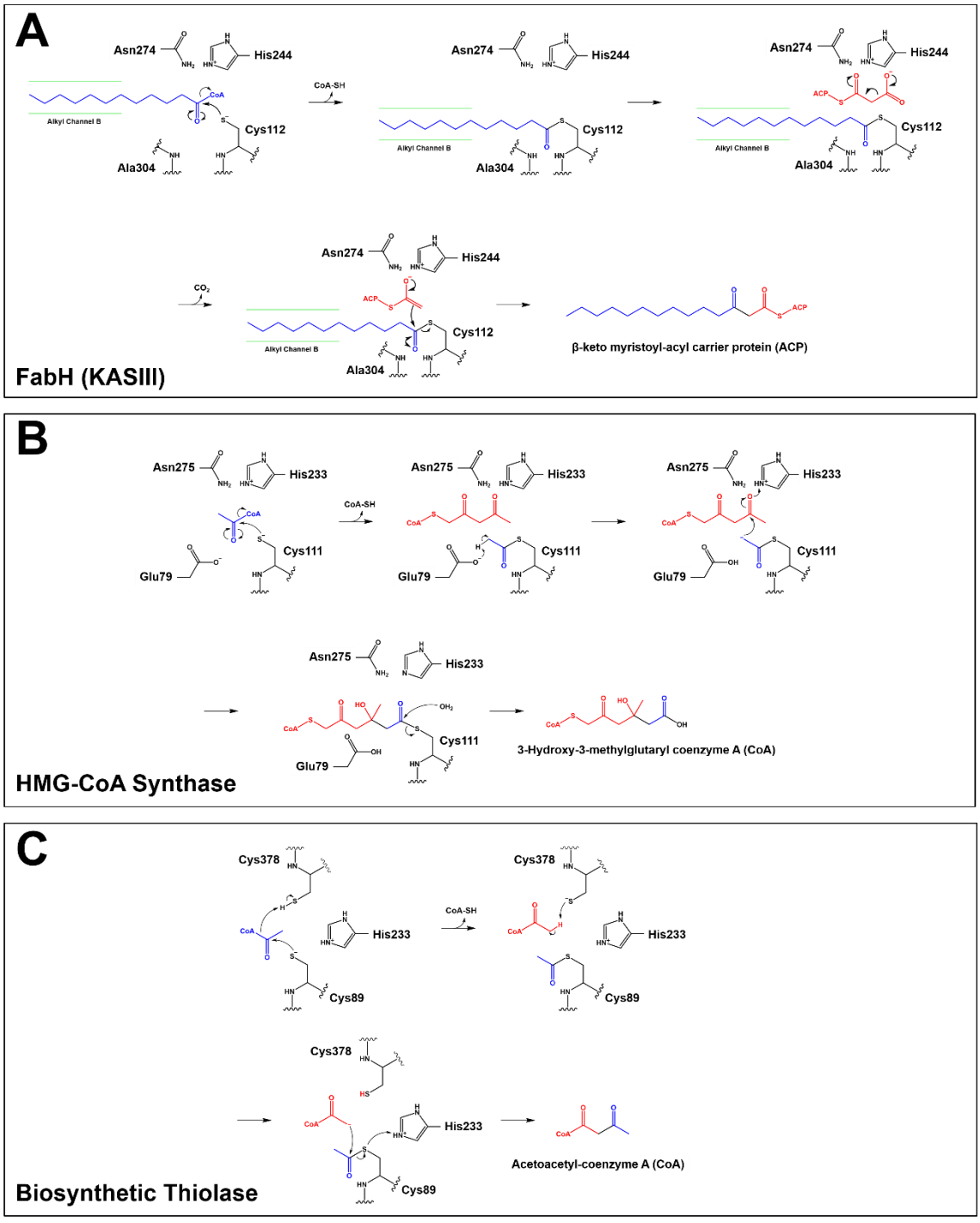


Figure 16: Comparison of mechanisms for FabH, HMG-CoA synthase, and biosynthetic thiolase enzymes. **A)** Mechanism of FabH *(21)*. Cys112 attacks the carbonyl of lauroyl-CoA (blue) bound in the OleA-equivalent alkyl channel B, releasing CoA-SH. Malonyl-ACP (red) binds in the active site pocket above this

acyl-enzyme intermediate and becomes decarboxylated. This nucleophile condenses the two substrates and releases product. **B)** Mechanism of HMG-CoA synthase *(28)*. Cys111 attacks the carbonyl of acetyl-CoA (blue), releasing CoA-SH. Glu79 abstracts a proton from C_2 of acetyl-Cys111, promoting condensation between nucleophilic acetyl-Cys111 and the second substrate, acetoacetyl-CoA (red). Hydrolysis by an activated water releases product. **C)** Mechanism of biosynthetic thiolase *(24)*. Cys89 attacks the carbonyl of acetyl-CoA (blue), releasing CoA-SH. Cys378 abstracts a proton from C_2 of the second substrate, acetoacetyl-CoA (red), promoting condensation between nucleophilic acetyl-CoA and acetyl-Cys89. Product is then released.

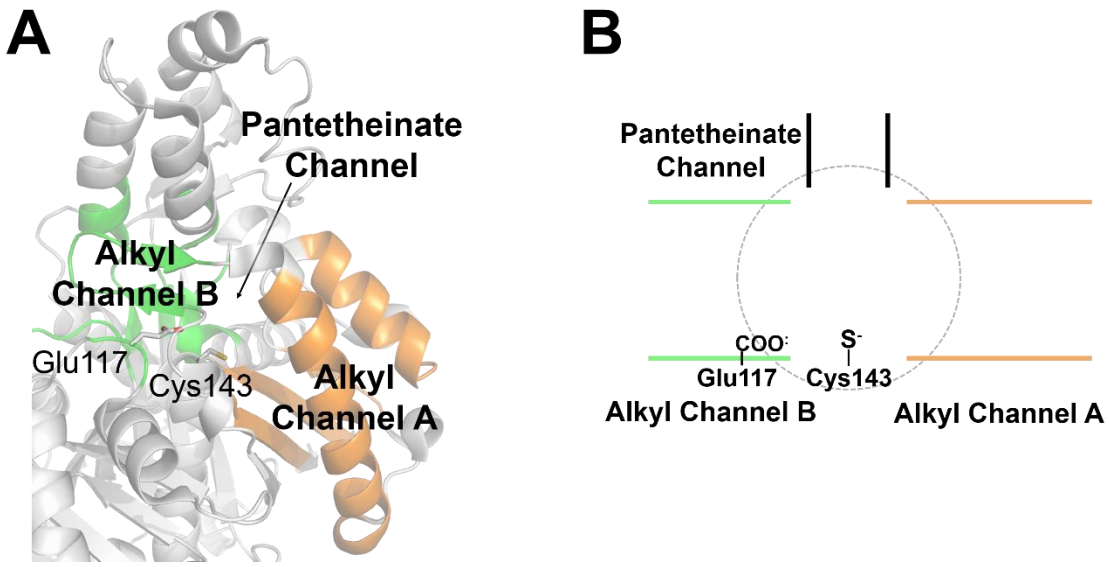


Figure 17: The three channels of the OleA active site. **A)** Monomer of WT OleA (PBD ID 3ROW) centered on Cys143, showing the structural elements that define alkyl channel A (orange) and alkyl channel B (green), with the solvent accessible pantetheinate channel indicated by a black arrow *(25,54)*. **B)** Cartoon representation of the three substrate binding channels centered on Cys143.

As the first committed step of olefin biosynthesis, OleA is an important target for protein engineering. *In vivo* studies have revealed that OleA controls the hydrocarbon produced by the Ole pathway (7). *Xanthomonas campestris* produces a wide variety of hydrocarbons ranging from C_{28} to C_{31} with one, two, or three double bonds *in vivo* (13). As a consequence, OleA from *X. campestris* is promiscuous and condenses fatty acyl-CoA substrates ranging from C_8 to C_{16} *in*

vitro (15). Therefore, the *X. campestris* Ole pathway represents a good target for protein engineering. Understanding the mechanism necessary to coordinate, activate, and condense long-chain hydrophobic substrates may provide insights needed to tailor the pathway for non-natural olefins.

In this study, we have solved the crystal structures of *X. campestris* OleA E117A, E117Q, and E117D mutant enzymes and cerulenin-inhibited states. These structures and the characterization of these enzymes suggest that Glu117 is responsible for initiating the condensation of the fatty acyl substrates during turnover by acting as a general base. Modeling studies suggest non-canonical substrate binding that contributes to diminished activity. To our knowledge, this is the first example of a thiolase superfamily enzyme using a catalytic base donated by the second subunit of the dimer.

Experimental

Preparation of OleA Glu117 site mutants

The initial cloning of the *Xanthomonas campestris* pv. *campestris* str. ATCC 33913 (NP_635607.1) synthetic *oleA* gene into pET28b^{+} (Novagen, Madison, WI) vector was previously described (15). Site-directed mutagenesis was performed using custom primers from Integrated DNA Technologies (IDT, Coralville, IA). Mutations to change the Glu117 residue in WT OleA to the corresponding E117A and E117Q were introduced via PCR using Phusion polymerase (New England Biolabs). Parental template DNA was removed by

DpnI digestion. The E117D variant was purchased as an *E. coli* codon-optimized gBlock (IDT) and was assembled into pET28b^{+} via Gibson assembly (New England Biolabs). Successful mutations were verified by DNA sequencing (ACGT, Inc.).

Expression and purification of OleA mutant enzymes

Mutant OleA was expressed in *Escherichia coli* BL21 (DE3) competent cells (Invitrogen). Flasks of 1 liter LB media containing 50 µg/mL kanamycin sulfate were inoculated with 5 mL saturated overnight culture carrying the appropriate *oleA* vector. Cells were grown at 37°C until an A_{600} of 0.5-0.6 was obtained. Expression was induced by addition of 100 µM isopropyl β-D-1-thiogalactopyranoside. After 4 hr at 37°C, cells were harvested by centrifugation at 4000 x *g* for 10 min.

Cells were resuspended in buffer consisting of 200 mM sodium chloride and 20 mM Tris HCl pH 7.4 containing a single protease inhibitor cocktail tablet (Roche Applied Science). Lysis was carried out using a sonicator (QSonica Q700). The soluble fraction was separated by centrifugation at 27,000 x *g* for 1 hr and passed through a 0.2 µm syringe filter. Filtered lysate was purified using nickel affinity chromatography (5 mL HisTrap HP column) driven by an ÄKTA FPLC (GE Healthcare Life Sciences). Protein was eluted using a linear gradient consisting of resuspension buffer with the addition of 500 mM imidazole pH 7.4. The purity and size of WT and mutant OleA were analyzed by SDS-PAGE and SimplyBlue SafeStain (Invitrogen). Fractions containing only OleA were pooled

and concentrated to below 2 mL. The sample was run over size exclusion chromatography using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare Life Sciences). The single peak at 80 min was collected, analyzed by SDS-PAGE, and concentrated to 115 µM (9 mg/mL) for crystallization trials. OleA E117A, E117Q, and E117D mutations were confirmed by mass spectrometry (Bruker Autoflex Speed Matrix-assisted laser desorption/ionization time of flight mass spectrometer, University of Minnesota Center for Mass Spectrometry and Proteomics).

Detection of coenzyme A hydrolysis by OleA mutants

The hydrolytic displacement of CoA from myristoyl-CoA by WT and mutant OleA enzymes were detected using 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) as described previously (54). OleA (20 µg) was incubated at room temperature with 65 µM myristoyl-CoA in 490 µL of 200 mM sodium chloride, 20 mM Tris HCl pH 7.4, and 4% v/v ethanol for incubation times ranging from 5 minutes to 24 hours. Reaction mixtures were quenched by denaturing the enzyme using rapid addition of 500 µL of 8.25 M guanidine HCl pH 7.4 before DTNB (300 µM) addition. After a ten minute incubation with DTNB, the free thiol, CoA-SH, was quantified spectrophotometrically at 412 nm (ε_{412}=13,600 $M^{-1}cm^{-1}$) (15,34,35). Enzyme-dependent production of CoA was controlled for by subtracting non-enzymatic hydrolysis contributions to absorbance for each mutant enzyme.

Detection of mutant OleA condensation of myristoyl-CoA substrate

WT and mutant OleA turnover assays were conducted in glass vials at room temperature. OleA (10 µg) was mixed with 50 µM myristoyl-CoA in buffer consisting of 200 mM sodium chloride and 20 mM Tris HCl pH 7.4. After incubations of 1, 3, and 24 hours, reactions were quenched with an equal volume of ethyl acetate. The organic phase was extracted and 5 µL was analyzed by gas chromatography, splitting the sample between the flame ionization detector and the mass spectrometer as previously described (56). Elution program conditions consisted of 250°C inlet temperature, 100°C start temperature, 10°C/min ramp rate, 320°C final temperature, and a 5 min hold time (total time of 27 min). The thermally-decarboxylated OleA product elution time was confirmed by the commercial standard, 14-heptacosanone (Sigma Aldrich).

pH dependence and apparent pK_a determination of WT and E117D OleA

The pH dependence of WT and E117D OleA activity was investigated by product condensation and myristoyl-CoA hydrolysis as described above using the following buffer conditions: 0.05 M sodium citrate (pH 5.5), 0.2 M sodium chloride; 0.05 M bis-tris propane (pH 6.0-9.5), 0.2 M sodium chloride; 0.05 M CAPS (pH 10.0-10.5), 0.2 M sodium chloride. WT OleA activity was measured in triplicate after 8 minutes, and E117D OleA activity was measured in triplicate after 24 hours. Nonlinear fitting analyses were performed using GraphPad Prism 7.03 (GraphPad Software, La Jolla California USA). WT OleA data exhibited sigmoidal behavior that rapidly diminished after pH 9.5 and were fit to the

equation $y = y_{lim}/(1+10^{pKa-pH})$ as described previously (57). As such, only the increase in activity going from acidic to basic conditions was compared. E117D OleA curves were fit to the Gaussian function in GraphPad Prism 7.03. Apparent pK_a values for WT were calculated as the inflection point of the curve, and apparent pK_a values for E117D were calculated as one standard deviation below the mean.

Crystallization of OleA Glu117 mutant enzymes

Initial identification of crystallization conditions for OleA was previously described (25). E117A, E117Q, and E117D protein crystals were grown under similar conditions utilizing the hanging-drop vapor diffusion method. Mutant-alone crystallizations produced crystals of poor diffraction quality, but similar morphology to WT OleA crystals. To improve diffraction quality and potentially trap catalytic intermediates, mutant OleA (115 µM, 9 mg/mL) was incubated with 1 mM myristoyl-CoA for times ranging from 1 to 24 hours. A 1 µL drop of substrate-incubated mutant OleA was combined with 1 µL of well solution (0.1 M sodium citrate pH 4.2, 0.10-0.12 M potassium phosphate dibasic, and 14-17% PEG 8000) and was suspended over 500 µL of well solution. Crystal trays were incubated at 19°C until crystals appeared (1-2 days). Both rod-shaped and bipyramidal crystals were harvested and cryoprotected in well solution containing 25% v/v glycerol before flash-freezing in liquid nitrogen.

Co-crystals of OleA with the fungal inhibitor cerulenin were previously reported (25). OleA E117A, E117Q, and E117D (115 µM, 9 mg/mL) were reacted

with 1 mM cerulenin for 1 hr. Dialysis was used to remove excess cerulenin and avoid phase separation in the crystallization drop. Successful crystallization conditions were as follows: 0.07-0.10 M manganese (II) chloride, 0.1 M MES pH 6.0. 13-15% w/v PEG 4000. Long, grooved crystals were harvested and cryoprotected in well solution containing 25% v/v glycerol before flash-freezing in liquid nitrogen.

X-ray data collection, processing, and refinement

All E117A and E117Q datasets were collected at GM/CA-CAT beamlines 23-ID-B (MARmosaic 300 CCD) and 23-ID-D (Pilatus3 6M) at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Datasets for E117D crystals were collected at SBC-CAT beamline 19-BM (ADSC Quantum 210r) at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Crystals were exposed to either 1.02 or 1.03 Å wavelength X-rays at 100 K with an attenuated beam (5-10-fold) set to match the size of each crystal. Strong electron density was observed at monomer crystal contacts for all mutant-cerulenin datasets, prompting the investigation of anomalous diffraction by a bound metal ion. As manganese (II) chloride was a major component of the crystallization well solution, anomalous diffraction data were collected for E117Q-cerulenin crystals above (1.85 Å) and below (1.94 Å) the manganese absorption edge. Recorded data were processed in either XDS or HKL2000 (58,59). The 1.85 Å resolution structure of the WT OleA structure (Protein Data Bank (PDB) ID 3ROW) was used as an initial phasing model to solve the isomorphous E117A,

E117Q, and E117D myristoyl-CoA co-crystal structures using direct Fourier synthesis in CCP4 (REFMAC5) (60,61). Although the same crystal form as WT OleA bound with cerulenin, the E117A-cerulenin co-crystal dataset required phasing by molecular replacement (PhaserMR) using the original 1.7 Å resolution WT OleA-cerulenin crystal structure (PDB ID 3S21) as a search model (62). This phasing method was used to assign the correct space group, $P3_221$, indicated by a strong increase in the translational *Z* score (TFZ) peak compared to $P3_121$ enantiomorphic space group (TFZ of 45.5 vs. 21.4, respectively). E117Q-cerulenin and E117D-cerulenin co-crystal structures were phased using the E117A-cerulenin structure by direct Fourier synthesis in REFMAC5. Reflections set aside for R_{free} were carried over from the WT OleA and WT OleA-cerulenin structures, respectively (25). Subsequent rounds of model building in COOT and restrained refinement in REFMAC5 were carried out for all structural models until all interpretable $2mF_o$-DF_c and mF_o-DF_c electron density was explained (63). When necessary, the occupancy of each alternative conformer for Cys143 was adjusted in 10% increments until the B factor values for the sulfur atoms were similar to each other and surrounding residues after refinement. A final round of restrained refinement with translation-libration-screw refinement was performed before statistical validation (MolProbity) (46). Root mean square deviation (rmsd) calculations were performed using the SUPER command in PyMOL, matching all alpha carbons between aligned structures (47).

Results

Mutagenesis strategy, expression, and purification of WT and mutant OleA enzymes

An amino acid sequence alignment of known OleA enzymes shows that Glu117 is invariant (Figure 18). In order to investigate the role of Glu117, three mutants were produced: Ala117, Gln117, and Asp117. The alanine mutation was chosen to probe the effect of no carboxylate group at the 117 position. Glutamine was selected to remove the acid/base properties but retain hydrogen bonding capabilities. Aspartate was introduced to reduce the length of the residue but retain the negative charge and the potential to act as a general base.

X. campestris	V	S	R	D	Y	L	^{117}**E**	P	S	T	A	S
G. uraniireducens	V	C	R	E	Q	F	^{97}**E**	P	A	T	A	C
S. oneidensis	V	C	R	D	Q	H	^{97}**E**	P	A	T	A	C
S. ambofaciens	T	S	Q	D	L	I	^{97}**E**	P	A	T	A	H
A. aurescens	V	T	R	R	N	L	^{103}**E**	P	S	V	A	V
K. radiotolerans	V	T	R	P	H	L	^{119}**E**	P	A	V	A	T
M. luteus	V	T	R	A	T	L	^{106}**E**	P	S	V	A	V
S. maltophilia	V	S	R	D	Y	L	^{97}**E**	P	S	T	A	S
C. aggregans	V	C	Q	D	Y	L	^{97}**E**	P	S	T	A	C

Figure 18: Amino acid sequence alignment of a subset of bacterial OleA enzymes showing the invariant glutamate residue. Bacterial genus and species names are as follows: *Xanthomonas campestris* pv. *campestris* str. ATCC 33913, *Geobacter uraniireducens* Rf4, *Shewanella oneidensis* MR-1, *Streptomyces ambofaciens* ATCC 23877, *Arthrobacter aurescens* TC1, *Kineococcus radiotolerans* SRS30216, *Micrococcus luteus* NCTC 2665, *Stenotrophomonas maltophilia* K279a, *Chloroflexus aggregans* DSM 9485.

WT and mutant OleA were recombinantly expressed and purified yielding between 15-20 mg OleA per liter of culture. Size exclusion chromatography produced a single, monodisperse population, corresponding to the OleA homodimer. The purity of this sample was estimated to be >95% pure by SDS-

PAGE. Mass spectrometry demonstrated that each OleA variant had the expected molecular mass.

Detection of coenzyme A hydrolysis by OleA mutants

The colorimetric DTNB assay measures free thiol groups produced either by hydrolysis through the transesterification and condensation steps of productive OleA catalysis or potentially futile off-pathway hydrolysis of the acyl-CoA substrates (Figure 19). Table 6 summarizes the results of WT and mutant OleA hydrolysis of myristoyl-CoA. WT OleA completely hydrolyzes available substrate in under 15 minutes. Only after 24 hour incubations with substrate are E117A and E117Q OleA mutants able to hydrolyze 25% of the substrate present, while E117D OleA hydrolyzes 50%.

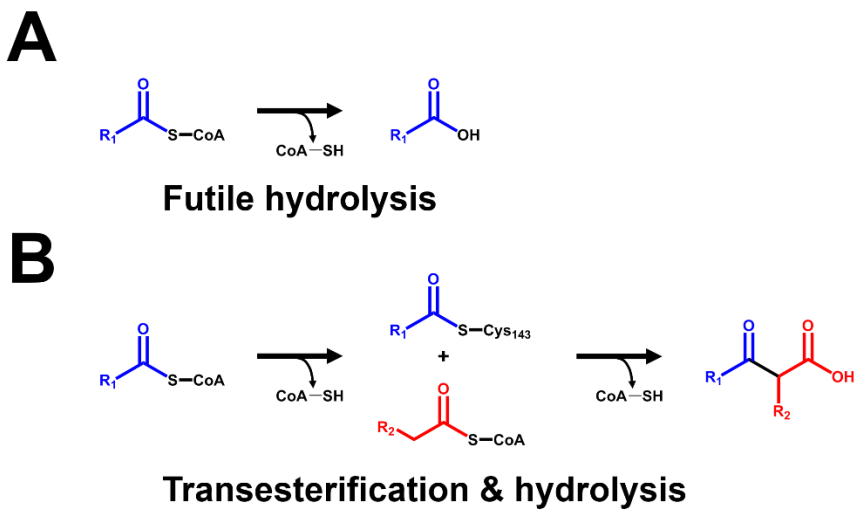


Figure 19: The OleA-catalyzed hydrolysis reactions with acyl-CoA substrate. **A)** Direct hydrolysis of acyl-CoA substrate by OleA, producing free CoA-SH and fatty acid. This reaction is designated as futile hydrolysis. **B)** Transesterification and condensation of acyl-CoA substrates by OleA, producing 2 moles of CoA-SH and β-keto acid.

Table 6: Hydrolysis of myristoyl-CoA by WT and mutant OleA enzymes.

Enzyme	Enzyme concentration (μM)	Rxn time (hrs)	CoA-SH product (nmol)[a]	Percent yield (%)[b]
WT	0.5	0.25	33 ± 0.5	100
E117A	0.5	0.25	0.8 ± 0.1	2.5
E117A	0.5	24	8.6 ± 0.3	27
E117Q	0.5	0.25	<0.1	<0.1
E117Q	0.5	24	7.8 ± 0.2	24
E117D	0.5	0.25	<0.1	<0.1
E117D	0.5	24	16 ± 0.9	50

Values shown are the average in triplicate with standard deviation.
[a] Free coenzyme A detected as described under "Experimental"
[b] Starting substrate was 32.5 nmoles; 32.5 nmoles product is 100% turnover

Detection of mutant OleA condensation of myristoyl-CoA substrate

The ability of OleA E117A, E117Q, and E117D mutants to form product was analyzed by gas chromatography/mass spectrometry (GC/MS). As the β-keto acid product is unstable in solution, the thermally-degraded ketone side product is observed in each chromatogram (Figure 20A). The commercially available ketone (14-heptacosanone) standard elutes at 16 minutes. The WT OleA product elutes at the same retention time and has an identical mass spectrum as the ketone standard (Figure 20B). After 15 minutes, WT OleA converts nearly 90% of substrate to product (Table 7). E117A and E117Q OleA exhibit no product formation discernible above baseline, even after 24 hours (Table 7) (Figure 20D). Only E117D OleA is able to condense substrate and in 24 hours produce the expected 14-heptacosanone product, albeit only ~25% conversion (Table 7) (Figure 20C).

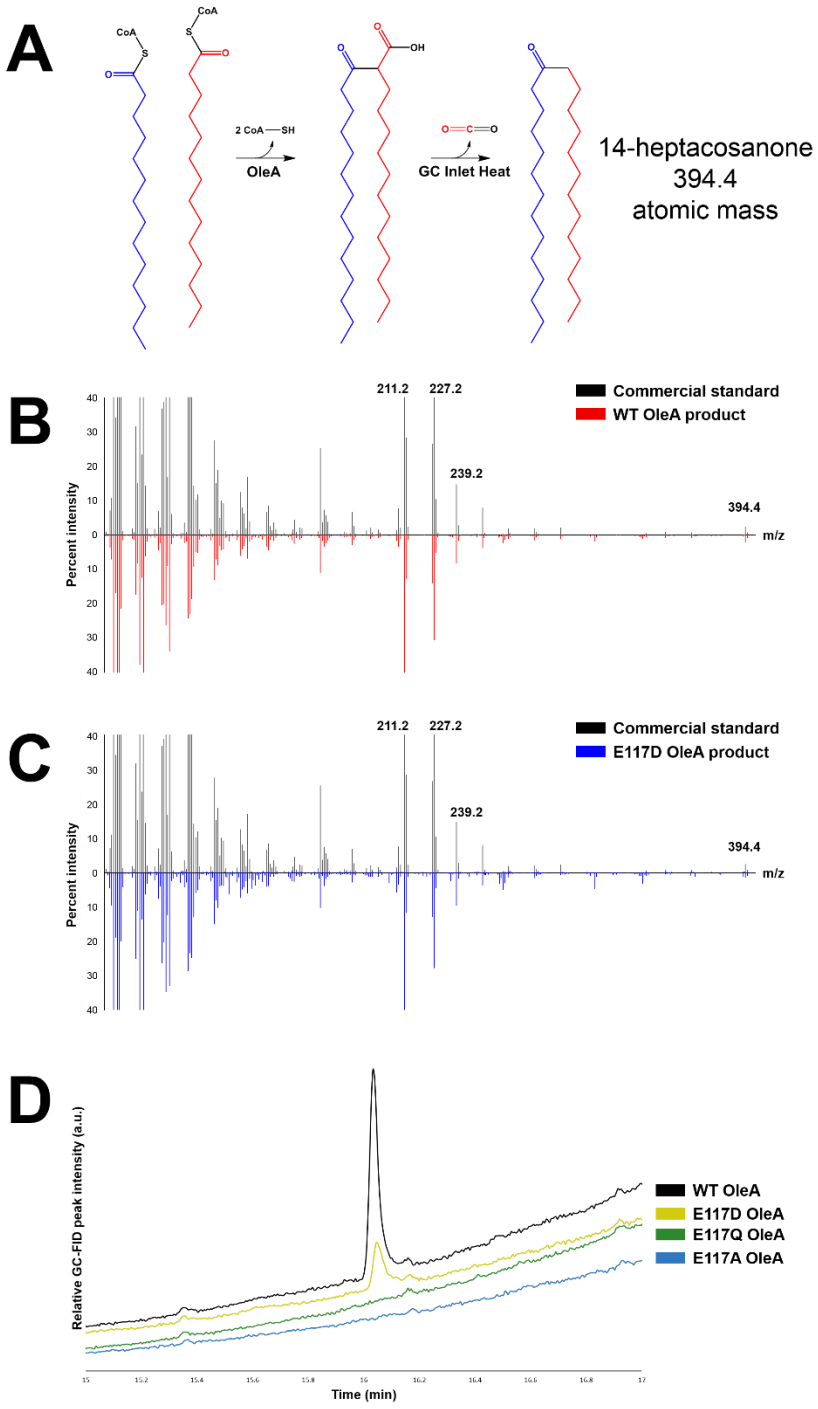


Figure 20: Confirmation of the ketone product of OleA WT and mutant turnover assay. **A)** OleA-catalyzed condensation of two myristoyl-CoA substrates to form β-keto acid. Thermal decarboxylation of this product to 14-heptacosanone (ketone) occurs during sample injection during gas chromatography. **B)** Mass spectral fingerprints (positive ion signal, expressed as percentage of highest

intensity) of the commercial standard 14-heptacosanone (black) and WT OleA catalyzed product (red). **C)** Mass spectral fingerprints (positive ion signal, expressed as percentage of highest intensity) of the commercial standard 14-heptacosanone (black) and E117D OleA catalyzed product (blue). **D)** GC-FID chromatograms for the product of WT (black), E117D (yellow), E117Q (green), and E117A (blue) OleA. The peak at 16 minutes represents 14-heptacosanone.

Table 7: Product turnover by WT and mutant OleA enzymes.

Enzyme	Enzyme concentration (μM)	Rxn time (hrs)	Ketone peak area (a.u.)[a]	Normalized ketone peak area (%)[b]
WT	0.25	0.25	5599 ± 146	86.1
WT	0.25	24	6510 ± 156	100
E117A	0.25	24	N.D.[c]	N.D.[c]
E117Q	0.25	24	N.D.[c]	N.D.[c]
E117D	0.25	24	1610 ± 131	24.7

[a]Values shown are the average in triplicate of the GC-FID peak area (arbitrary units) with standard deviation.
[b]GC-FID peak area as a percentage of WT 24 hour reaction peak area
[c]N.D: not detectable; peak area below 50 a.u. considered indistinguishable from baseline

pH dependence and apparent pK_a determination of WT and E117D OleA

WT and E117D OleA product condensation and myristoyl-CoA hydrolysis activities were tested at varying pH levels in order to determine an apparent pK_a for the ionized residue involved in each assay (Figure 21). General behavior towards pH was identical across both assays for each enzyme, but different between WT and E117D OleA. Although a bell shaped curve was obtained for E117D OleA, WT exhibited sigmoidal pH dependency as activity increased from acidic to basic pH, but a precipitous loss of activity was observed above pH 9.5. As such, we assigned the pKa of residue 117 to the acidic pKa of each assay, where behavior was similar for the two enzymes. For myristoyl-CoA hydrolysis, the pK_a for WT OleA (6.4) is ~1 pH unit lower than the E117D mutant enzyme (7.2) (Figure 21A). A similar trend is observed for β-keto acid production, showing pK_a values of 6.9 for WT and 7.5 for E117D OleA (Figure 21B). A

general shift from acidic to more basic pK_a values from myristoyl-CoA hydrolysis to product formation experiments is present for both OleA enzymes.

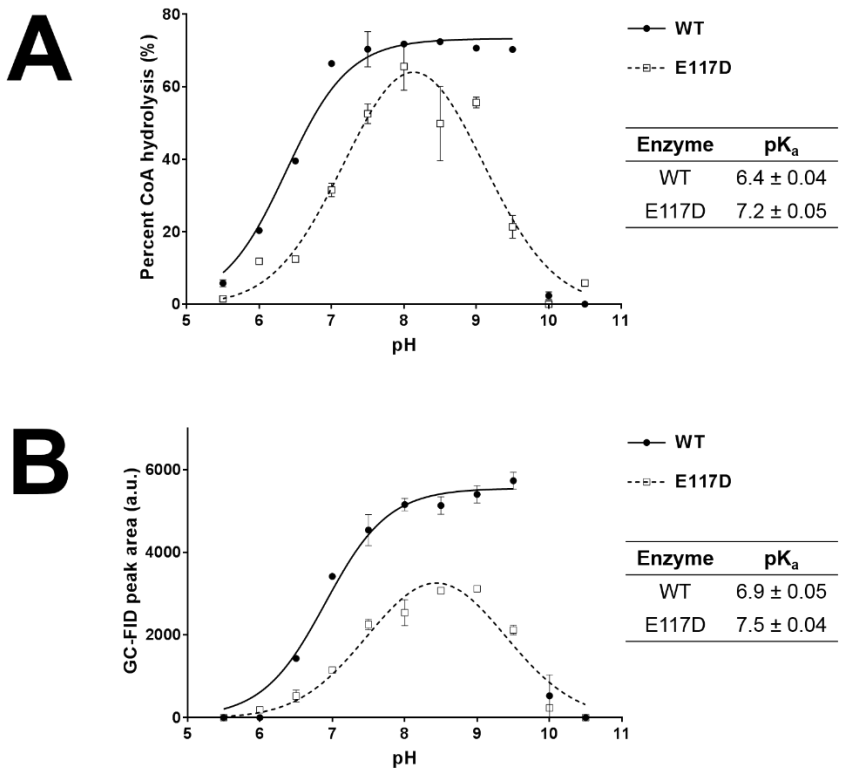


Figure 21: OleA WT and E117D enzyme activity vs. pH curves. **A)** Myristoyl-CoA hydrolysis by WT (black circles) and E117D (white squares) OleA at indicated pH levels. Calculated pK_a values for each enzyme are shown in the inset table. **B)** Turnover of β-keto acid by WT (black circles) and E117D (white squares) OleA at indicated pH levels. Calculated pK_a values for each enzyme are shown in the inset table. All WT OleA activity was measured after eight minutes, and all E117D OleA activity was measured after 24 hours. Results are presented as the mean of triplicate data ± standard deviation. pK_a values were calculated from nonlinear regression of each curve as detailed in "Experimental" and are presented in each table ± standard error for that variable.

Crystallization, data processing, and refinement of OleA Glu117 mutant structures

Data collection and refinement statistics for all mutants and cerulenin co-crystals are reported in Table 8. The mutant-alone E117A, E117Q, and E117D OleA structures were solved to 1.95 Å, 1.75 Å, and 1.84 Å resolution, respectively. All were solved in the $P2_12_12_1$ space group with one OleA homodimer in the asymmetric unit (Table 8). The rmsd over all alpha carbon atoms between WT OleA and E117A, E117Q, and E117D mutants are 0.23 Å, 0.17 Å, and 0.13 Å, respectively. The E117A-cerulenin, E117Q-cerulenin, and E117D-cerulenin structures were solved to 1.66 Å, 2.07 Å, and 2.08 Å resolution, respectively. As with WT-cerulenin, the mutant-cerulenin complexes were in the $P3_221$ space group with a single monomer in the asymmetric unit (Table 8). A crystallographic symmetry mate completes the dimeric biological assembly. The calculated rmsd values between OleA WT-cerulenin and E117A-cerulenin, E117Q-cerulenin, and E117D-cerulenin are 0.15 Å, 0.15 Å, and 0.17 Å, respectively. The $P3_221$ crystal form was only attained with cerulenin binding and could not be achieved with unbound or substrate reacted forms. Similar to WT OleA-cerulenin, each mutant-cerulenin complex contains a metal ion that is present at a crystal contact. This ion was confirmed to be a manganese ion from the crystallization by anomalous diffraction (Figure 22A).

Table 8: OleA Glu117 mutant data collection and refinement statistics[a].

	E117A OleA	E117Q OleA	E117D OleA	E117A OleA–Cerulenin cocrystal	E117Q OleA–Cerulenin cocrystal	E117D OleA–Cerulenin cocrystal
Data Collection						
Wavelength (Å)	1.03	1.03	1.02	1.03	1.03	1.02
Space group	P $2_1 2_1 2_1$	P $2_1 2_1 2_1$	P $2_1 2_1 2_1$	P 3_2 2 1	P 3_2 2 1	P 3_2 2 1
Unit cell (Å)	82.4 x 86.0 x 103.6	82.2 x 85.9 x 103.0	81.9 x 85.4 x 102.8	90.0 x 90.0 x 69.6	90.3 x 90.3 x 69.4	89.8 x 89.8 x 69.5
Unit cell (°)	α = β = γ = 90	α = β = γ = 90	α = β = γ = 90	α = β = 90; γ = 120	α = β = 90; γ = 120	α = β = 90; γ = 120
Resolution (Å)	50.0-1.96 (1.99-1.96)	50.0-1.75 (1.78-1.75)	50.0-1.84 (1.87-1.84)	50.0-1.66 (1.69-1.66)	29.6-2.07 (2.13-2.07)	50.0-2.08 (2.12-2.08)
Measured reflections	210319	544020	463599	217730	115512	206002
Unique reflections	52682	73521	63324	38595	20256	19937
Completeness (%)	99.9 (100)	99.9 (100)	100 (100)	99.3 (99.0)	99.7 (96.7)	98.8 (100)
R_{merge} (%)[b]	10.1 (71.4)	8.0 (82.0)	8.8 (97.5)	7.7 (86.5)	8.1 (80.4)	14.3 (72.0)
I/σ_I	11.8 (2.0)	10.8 (2.8)	11.7 (2.5)	10.7 (2.1)	17.6 (2.0)	11.8 (3.5)
CC*	0.971 (0.857)	0.987 (0.947)	0.986 (0.921)	0.982 (0.905)	0.999 (0.745)	0.988 (0.925)
Multiplicity	4.0 (4.0)	7.4 (7.4)	7.3 (6.9)	5.6 (5.6)	5.7 (4.9)	10.3 (10.0)
Wilson B factor (Å^2)	26.8	25.5	23.9	21.7	32.2	36.1
Source	APS 23 ID-B	APS 23 ID-B	APS 19-BM	APS 23 ID-B	APS 23 ID-B	APS 19-BM
Processing software	HKL2000	HKL2000	HKL2000	HKL2000	XDS	HKL2000
Refinement						
Resolution (Å)	1.96	1.75	1.84	1.66	2.07	2.08
Monomers in ASU	2	2	2	1	1	1
R_{work}[c]	0.1688	0.1612	0.1582	0.1741	0.1768	0.1711
R_{free}[d]	0.2141	0.1974	0.1979	0.2096	0.2443	0.2289
Ramachandran statistics (%)[e]						
favored	96.18	96.0	96.2	96.3	96.0	96.7
outliers	0.44	0.60	0.30	0.00	0.31	0.30
RMS deviation						
bond lengths (Å)	0.0266	0.0292	0.0267	0.0314	0.0214	0.0246
bond angles (°)	2.410	2.511	2.448	2.757	2.098	2.306
Average B-factor (Å^2)	33.9	31.7	28.9	34.3	44.0	48.5
ligand B-factor (Å^2)	–	–	–	30.0	49.2	67.2
Clashscore (percentile)[e]	3.68 (99th)	3.58 (98th)	3.72 (98th)	8.40 (79th)	5.59 (97th)	5.44 (97th)
Protein Data Bank ID	5VXD	5VXF	5VXH	5VXE	5VXG	5VXI

[a]Data in parentheses are for the highest resolution shell. [b]$R_{merge} = \Sigma_i |I_{hkl,\,i} - \{I_{hkl}\}| / \Sigma_{hkl} \Sigma_i I_{hkl,\,i}$, where I is the observed intensity and $\{I_{hkl}\}$ is the average intensity of multiple measurements. [c]$R_{work} = \Sigma||F_o|-|F_c|| / \Sigma|F_o|$, where $|F_o|$ is the observed structure factor amplitude and $|F_c|$ is the calculated structure factor amplitude. [d]R_{free} is the R factor based on 5% of the data excluded from refinement. [e]Based on values obtained from MolProbity (46).

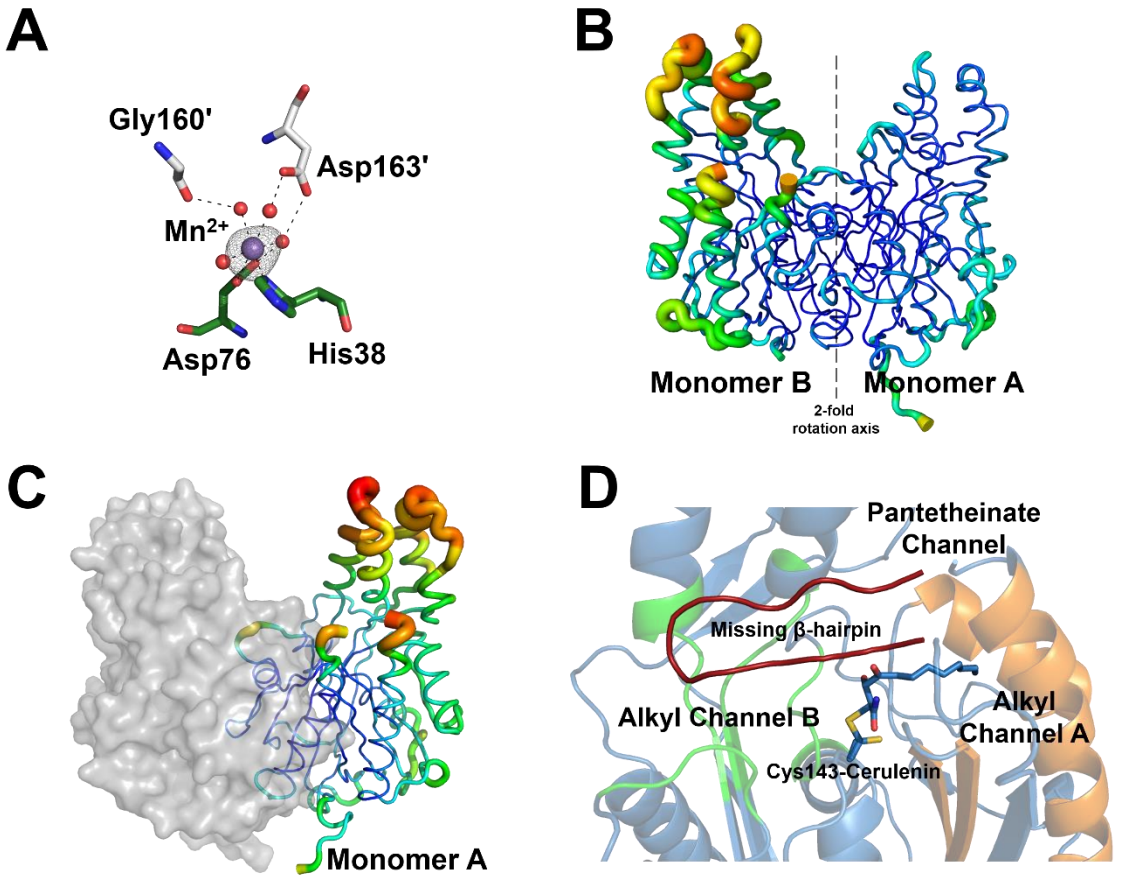


Figure 22: Structural features of OleA Glu117 mutants. **A)** OleA E117Q-cerulenin showing manganese ion coordination at the crystal contact between chain A (green) and the symmetry mate (gray). The anomalous map is contoured at 3σ. **B)** OleA E117D structure modelled by low (blue) to high (red) B factor magnitude. Monomer B shows the most disorder. **C)** OleA E117D-cerulenin bound structure represented by low (blue) to high (red) B factor magnitude. The crystallographic symmetry-related monomer completing the dimer is shown as a gray surface model. **D)** OleA E117A-cerulenin complex (blue, monomer A) showing the position of the mobile β-hairpin (red) over alkyl channel B that is missing from the electron density.

Although co-crystallized with myristoyl-CoA, the E117A, E117Q, and E117D OleA are presented with no substrate, product, or derivative bound and are reported as unbound, mutant-alone structures. Data from crystals grown in the absence of myristoyl-CoA were of inferior diffraction quality to crystallization

in the presence of substrate and were not pursued. Additional attempts to co-crystallize mutants with C_8-C_{12} and C_{16} acyl-CoA substrates, or to soak mutant crystals with substrate, led to empty active sites and inferior diffraction. Cerulenin co-crystallization experiments were performed to assess covalent binding capacity of each mutant enzyme and how each mutant positions covalent adducts.

Overall structures of OleA Glu117 mutants

Previous OleA studies have demonstrated that three substrate-binding channels emanate from the active site centered on the catalytic Cys143 (
Figure **17**) (25,54). All three Glu117 mutant OleA structures show these channels with no major changes to their size or shape. Aligning the active site residues of all three mutant enzymes with WT OleA shows little perturbation of active site residue positions in both monomer active sites. However, multiple conformers and side chain rotamers are observed in the E117A and E117Q OleA structures compared to WT (Figure 23). Notably, in the E117A and E117Q structures, Cys143 adopts two conformers in rapid rotational equilibrium differing by ~120° (Figure 23B & C). The occupancy of each conformer is 60%:40% for the E117A structure and 30%:70% for the E117Q structure (toward residue 117:away from residue 117). The E117D mutant contains only the orientation of Cys143 observed in WT, which is positioned away from Asp117 (Figure 23D). This may suggest the orientation of Cys143 is influenced by negative charge repulsion between the thiolate of Cys143 and the carboxylate of Glu117. Each Glu117

mutant contains an active site water molecule bound in the second oxyanion hole formed by the main-chain amides of Cys143 and Ser347. The E117D structure contains a second water molecule held in place by hydrogen bonds between the side chain carboxyl of Asp117, the side chain hydroxyl of Ser347, and the water molecule bound in oxyanion hole 2 (Figure 23D). Oxyanion hole 1, formed by the side chains of His285 and Asn315, indicates a lack of ordered water at this site in all three structures. Compared to WT, the side chain hydroxyl group of Ser347 is rotated away from Cys143 in the E117A and E117Q OleA structures (Figure 23B & C). Only in the WT and E117D OleA structure does the side chain of Ser347 rotate toward the catalytic cysteine, forming a hydrogen bond with a water molecule in E117D (Figure 23A & D). However, the Asp side chain is too short to form an additional hydrogen bond to the side chain of Ser347, as seen in WT. All three structures exhibit increased disorder in monomer B of the homodimer, with monomer A showing more order due to extensive crystal contacts (Figure 22B).

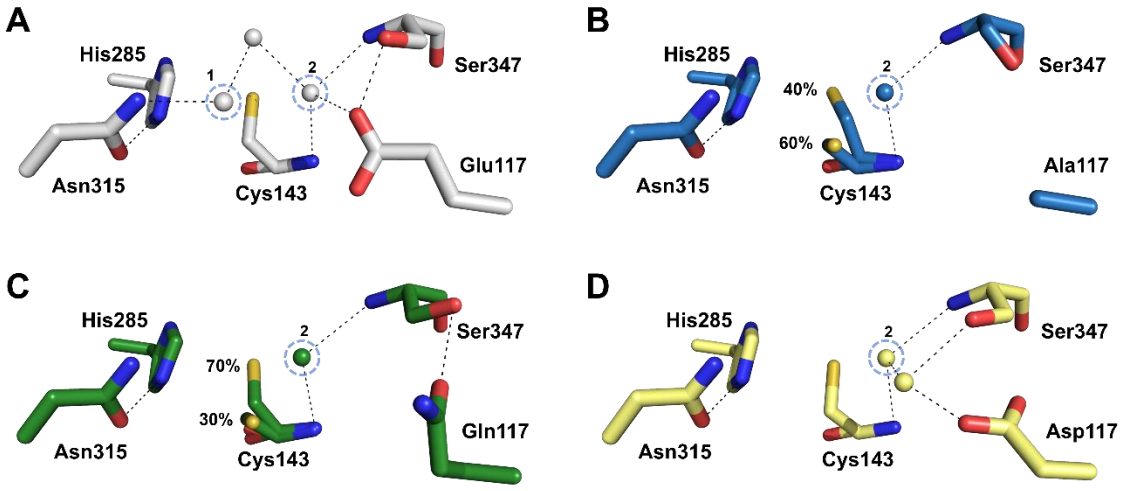


Figure 23: OleA WT and mutant active site comparison. **A)** Active site of WT OleA (PDB ID 3ROW) showing the hydrogen bonding network and water molecules (gray) bound in both oxyanion holes (1 and 2, blue dashed circles) (25). **B)** Active site of OleA E117A showing the hydrogen bonding network and a water molecule (blue) bound in the second oxyanion hole (2, blue dashed circle). Cys143 adopts two conformers at 60%:40% occupancy. **C)** Active site of OleA E117Q showing the hydrogen bonding network and a water molecule (green) bound in the second oxyanion hole (2, blue dashed circle). Cys143 adopts two conformers at 70%:30% occupancy. **D)** Active site of OleA E117D showing the hydrogen bonding network and water molecules (yellow). All hydrogen bonds are represented by black dashed lines. Monomer A was used to generate each figure.

As these crystal forms were co-crystallized with myristoyl-CoA substrate, several possible ligands (substrate, product, and intermediates), orientations, and occupancy levels were modeled into both alkyl channels A and B in order to test for low occupancy binding that was suggested by spurious broken density. Modelling myristoyl-CoA, myristic acid, and the acyl-Cys143 intermediate within alkyl channel A all produced maps lacking ligand density ($2mF_o$-DF_c) and/or positive difference density (mF_o-DF_c) in this channel following refinement. These models also caused an increase in R_{free} during refinement. Only the modelling

and refinement of myristoyl-CoA in alkyl channel B at 10% occupancy caused a significant increase in connected positive mF_o-DF_c map for the entire ligand and decreased R_{free}. This also revealed positive difference electron density for a second alkyl chain conformer within alkyl channel B, and vice versa when only this second conformer was modelled. Modeling and refining both conformers of myristoyl-CoA simultaneously into alkyl channel B at 10% occupancy each still led to significant positive mF_o-DF_c difference density for the full length of ligand (Figure 24). However, simulated annealing omit maps did not reveal evidence of bound ligand and further adjustments to occupancy were unsuccessful in fully explaining the map with no observable improvements to the $2mF_o$-DF_c electron density. The indication of low level binding of substrate is only observed in alkyl channel B of monomer B in the E117A and E117Q OleA mutant structures, which is consistent with previous reports of substrate binding only in monomer B of the dimer due to less extensive crystal contacts (54). These observations support that the electron density is not solely due to model bias, but ultimately no ligand was included during final refinement of the E117A and E117Q OleA structures.

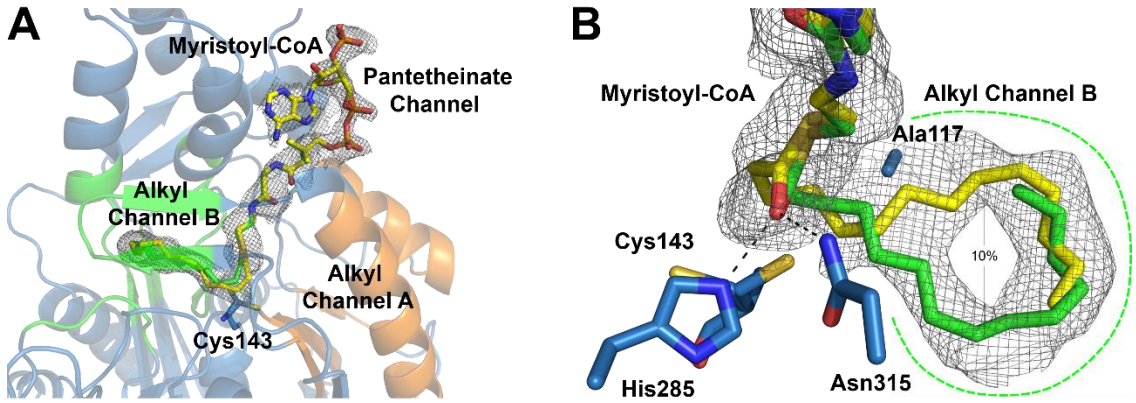


Figure 24: Low-occupancy, non-canonical binding of myristoyl-CoA in alkyl channel B of OleA E117A. **A)** Binding of myristoyl-CoA substrate (stick model) bound in alkyl channel B (green cartoon) of monomer B as seen from the orientation in Figure S5. Positive mF_o-DF_c map is contoured at 3σ. **B)** 180° rotation of myristoyl-CoA bound in alkyl channel B relative to panel A, viewed looking down the pantetheinate channel. The thioester carbonyl of the substrate is stabilized by the side chains of His285 and Asn315 (oxyanion hole 1). Each alternative conformer of myristoyl-CoA is modelled at 10% occupancy and occupies a different orientation within alkyl channel B. Electron density for the alternative alkyl conformers appeared in the absence of one model, but the presence of the other. Occupancy was so low that substrate binding was unclear in simulated-annealing omit maps, and so final refinement and PDB deposition did not contain the model.

Cerulenin-bound structures of OleA Glu117 mutants

OleA Glu117 mutants were covalently modified with cerulenin before crystallization. Cerulenin, a fungal inhibitor of fatty acid biosynthesis, acts to covalently modify active site cysteine (OleA Cys143) residues through reaction with its epoxide ring (Figure 25) (64,65). E117A, E117Q, and E117D OleA have refined occupancies of bound cerulenin at 70%, 70%, and 100%, respectively (Figure 26). In the E117A and E117Q OleA structures, a second, unbound conformer of Cys143 is present at 30% occupancy (Figure 26B & C). Cerulenin is bound exclusively in alkyl channel A as observed in the WT OleA-cerulenin

structure (25). The electron density indicates that the alkyl chain of cerulenin exhibits increased mobility as it extends deeper into the binding channel. The C3 hydroxyl of cerulenin occupies oxyanion hole 1 while the oxygen of the carboxamide moiety occupies oxyanion hole 2. The side chains of Ser347 and Gln117 form hydrogen bonds with the amine of the cerulenin carboxamide. Similar to WT OleA, Asp117 forms a hydrogen bond with Ser347 (Figure 26A & D). As with WT, all three cerulenin-bound mutant structures exhibit increased disorder including a lack of clear electron density for the β-hairpin that closes alkyl channel B from solvent (Figure 22C & D).

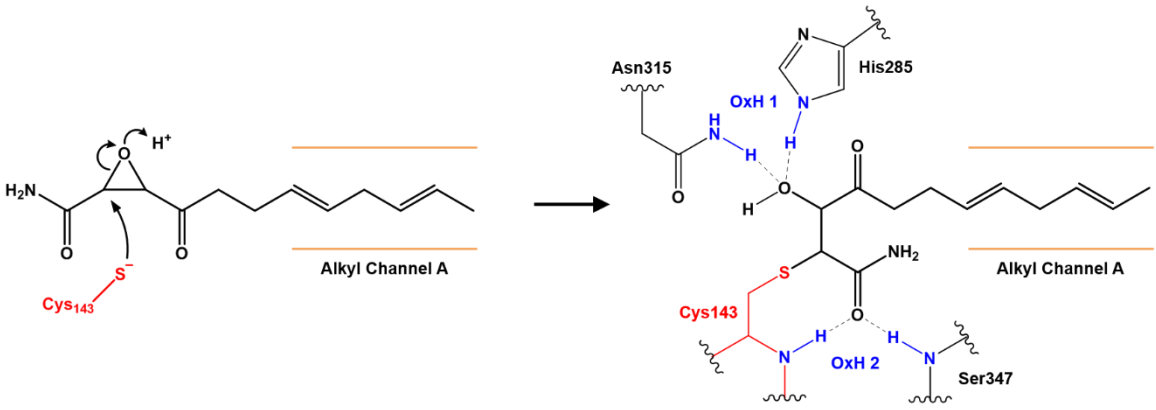


Figure 25: Covalent inhibition of OleA Cys143 by cerulenin. Upon binding, cerulenin occupies alkyl channel A in WT OleA (PDB ID 3ROW) and is stabilized by interactions with both oxyanion holes (OxH, blue) (25). Oxyanion hole 1 is formed by the side chains of His285 and Asn315. Oxyanion hole 2 is formed by the main-chain amides of Cys143 and Ser347. Dashed black lines represent hydrogen bonds.

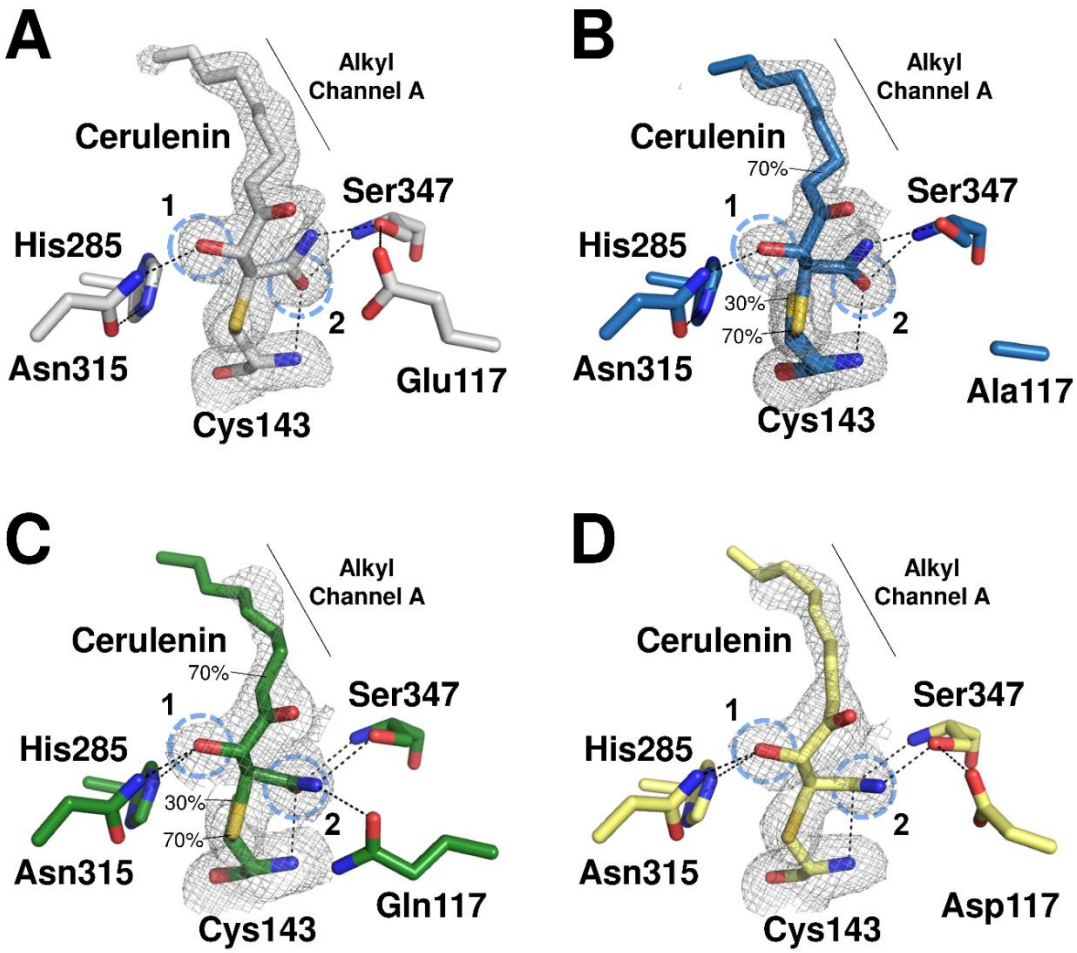


Figure 26: Cerulenin-bound mutant OleA active site. **A)** Active site of WT OleA-cerulenin (gray, PDB ID 3S21) with ligand bound in alkyl channel A at 100% occupancy (25). **B)** Active site of OleA E117A-cerulenin (blue) with ligand bound in alkyl channel A at 70% occupancy. **C)** Active site of OleA E117Q-cerulenin (green) with ligand bound in alkyl channel A at 70% occupancy. **D)** Active site of OleA E117D-cerulenin (yellow) with ligand bound in alkyl channel A at 100% occupancy. All hydrogen bonds are represented as black dashed lines. In all panels, oxyanion hole 1 (1, blue dashed circle) is occupied by the C3 hydroxyl of cerulenin, and oxyanion hole 2 (2, blue dashed circle) is occupied by carboxamide carbonyl oxygen of cerulenin. The simulated annealing ligand omit mF_o-DF_c maps are contoured at 2σ.

Discussion

In order to further understand the mechanism of OleA and its subcategory classification in the thiolase superfamily of enzymes, we solved the structures of

the catalytically-impaired OleA E117A, E117Q, and E117D mutant enzymes. Glu117 is absolutely conserved across OleA sequences and is located on a loop that in all mutants and WT is inserted into the opposite monomer active site of the physiological homodimer (Figure 27). Due to the proximity of Glu117 to bound substrates trapped in Cys143 mutant enzyme crystal structures, it was proposed that Glu117 acts as the general base needed to prime condensation between two fatty acyl-CoA substrates (54). These structures steered us to investigate the impact Glu117 mutants have on turnover, structure, and ability to bind substrates, leading us to present an updated mechanism for OleA (Figure 28).

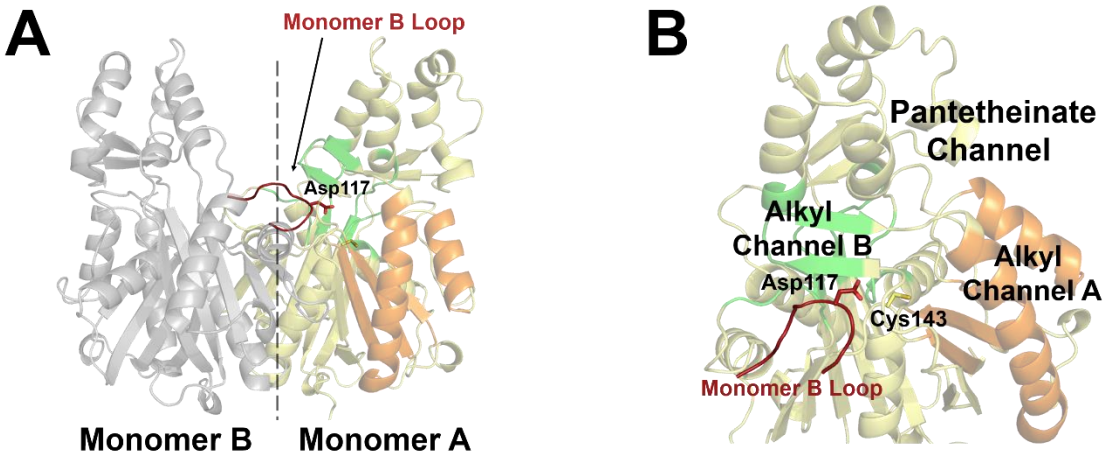


Figure 27: OleA E117D showing the dimer interface. **A)** OleA E117D dimer highlighting the insertion of the monomer B loop (red) containing Asp117 into the active site of monomer A. **B)** Monomer A of OleA E117D showing the active site loop containing Asp117 donated by monomer B (red).

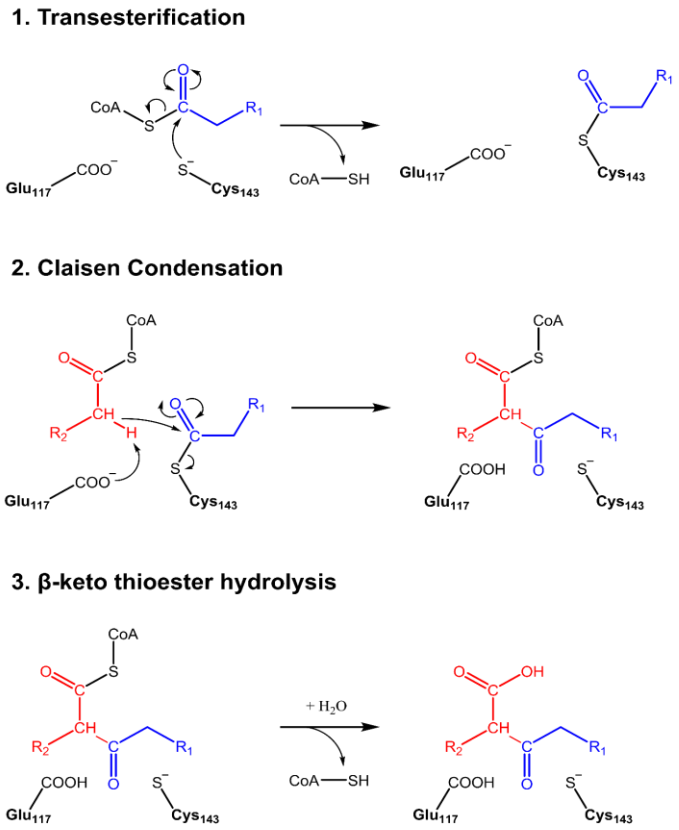


Figure 28: Proposed three-step mechanism of β-keto acid formation by OleA. Glu117 acts as a general base that deprotonates C_2 of the second acyl-CoA substrate to initiate condensation.

The first step in the mechanism of OleA involves the transesterification of the acyl moiety of the first acyl-CoA substrate. The only proposed role for Glu117 had been to initiate Claisen condensation (54). However, it is clear that all Glu117 mutants can catalyze CoA-SH release, as is evident by decreased rates of myristoyl-CoA hydrolysis and a shift in the pK_a of the pH dependence curve for E117D OleA (Table 6) (Figure 21A). E117A and E117Q OleA are only able to hydrolyze ~25% of the available myristoyl-CoA after a 24 hour incubation, whereas WT OleA has completely hydrolyzed this amount of myristoyl-CoA in under 15 minutes (Table 6). This demonstrates that these mutants are

hydrolyzing multiple myristoyl-CoA, and if this residue is absolutely required to act as a general base for condensation, a process of catalyzed hydrolysis is still occurring and the enzyme is not trapped within the first turnover. The ability of E117D OleA to release CoA-SH is also impacted even though it retains a carboxylate group in the active site similar to WT and can form β-keto acid. The apparent pK_a of the E117D mutant compared to WT OleA is shifted to a more basic pK_a by 0.8 pH units consistent with the loss of the hydrogen bond to Ser347 (Figure 23A & D). The pK_a values for the WT and E117D myristoyl-CoA hydrolysis reactions are more acidic than the pK_a values for product formation, indicating a less hydrophobic environment, which is more consistent with this pK_a being associated with CoA-SH release when only a single myristoyl-CoA is bound (Figure 21). Release of CoA-SH from the second substrate following transesterification would occur in a more hydrophobic environment in which two alkyl chains would be present in the active site. The OleA Glu117 mutant levels of myristoyl-CoA hydrolysis are higher than the previously reported C143A and C143S OleA mutants that lack the ability to undergo transesterification (54). This suggests OleA E117A and E117Q mutants may still be able to form low levels of acyl-enzyme intermediate, albeit at a diminished rate, and the reactivity with cerulenin confirms Cys143 can still act as a nucleophile (Figure 26B & C). Our attempts to trap the acyl-enzyme intermediate by co-crystallization with these mutants were unsuccessful indicating that this species, if present, does not accumulate. However, it should be noted that modelling of substrate into the

electron density of these mutants indicated low level binding of substrate in alkyl channel B, which is where the second substrate binds in WT OleA following transesterification. As such, the higher rate of hydrolysis could also be a consequence of an increased rate of futile hydrolysis when the first substrate incorrectly occupies alkyl channel B rather than alkyl channel A (Figures 19 & 24). Based on active site architecture, hydrolysis is likely mediated by an activated water with a candidate being the water molecule bound in oxyanion hole 2 (Figure 23B, C, & D). As E117A and E117Q OleA cannot form product, the myristoyl-CoA hydrolysis is slow and futile, releasing free CoA-SH and fatty acid (Figure 19A). In the case of E117D OleA, this must be able to catalyze transesterification as it can form β-keto acid. The environmental changes to active site pK_a and hydrogen bonding capabilities these mutations introduce likely play a role in decreasing the rate of myristoyl-CoA hydrolysis, whether it occurs through transesterification or direct hydrolysis by an activated water molecule.

The second step of the OleA mechanism is the Claisen condensation reaction (Figure 28). Production of β-keto acid only occurs if a new carbon-carbon bond forms between the two alkyl substrates. E117A and E117Q OleA are unable to condense substrate, indicating that Claisen condensation cannot proceed without an acidic residue at position 117 (Table 7). In line with Glu117 acting as the general base, E117D OleA is the only tested mutant able to produce β-keto acid. As with WT OleA, Asp117 is able to act as a base and is

capable of forming a hydrogen bond with the hydroxyl group of Ser347 when cerulenin is bound in alkyl channel A (Figure 26D) (25). This interaction has been suggested to be important for positioning and stabilizing Glu117 for proton abstraction, likely by helping define the pK_a of the carboxylate (54). Since this hydrogen bond can still form in the OleA E117D-cerulenin mutant, the diminished rate of product formation likely comes from the shorter length of aspartate compared to the native glutamate leading to sub-optimal positioning of the side chain for proton abstraction. In addition to slowing the rate of β-keto acid formation, E117D OleA causes a shift in pK_a compared to WT (Figure 21B). As the pK_as observed for both WT and E117D OleA (6.9 and 7.5, respectively) are different from those observed for myristoyl-CoA release (6.4 and 7.2, respectively), this is consistent with the pK_as being associated with a different chemical step, likely the condensation step of the OleA mechanism. The E117D mutant causes a 0.6 pH unit increase in the pK_a (6.9 to 7.5), indicating that Glu117 is involved in this step (Figure 21B). Structural changes caused by this mutation cannot be ruled out as a cause for this shift in apparent pK_a. However, OleA WT and E117D structures show very few structural differences (0.13 Å rmsd), lending further evidence to the association of this pK_a with Claisen condensation. Compared to the pK_a values for myristoyl-CoA hydrolysis, the pK_a values for OleA WT and E117D substrate condensation reactions are more basic, indicating a more hydrophobic environment. This is consistent with two bound alkyl chains in the active site, as would occur prior to condensation. The

increased hydrophobicity due to the second substrate binding and from which we propose Glu117 abstracts a proton, makes ionization in the active site less favorable, causing an increase in the apparent pK_a for Glu117.

The potential for non-canonical binding of substrate may affect the rates of transesterification and Claisen condensation in OleA Glu117 mutants. It has been proposed that during turnover, the first substrate binds in alkyl channel A for transesterification. This is based on cerulenin binding in alkyl channel A of OleA, as also observed in the Glu117 mutants, and the positioning of substrate trapped in crystals of C143A and C143S OleA (Figure 26) (25,54). Therefore, it is surprising that there is evidence of myristoyl-CoA in alkyl channel B in the OleA E117A and E117Q structures, but not in alkyl channel A (Figure 24). If OleA E117A and E117Q mutants improperly bind substrate in alkyl channel B, reduction in hydrolysis and turnover activity compared to WT would be expected. Binding in alkyl channel B exposes substrate to a different set of residues than is found at the opening to alkyl channel A, including the mobile β-hairpin that sequesters alkyl channel B from solvent. The increased flexibility and different chemical environment in alkyl channel B may facilitate futile hydrolysis of substrate in these mutants. Hydrolysis in alkyl channel B would allow for release of myristic acid from the active site due to dynamics of the β-hairpin, causing an apparent increase in myristoyl-CoA hydrolysis compared to that of catalytically dead Cys143 mutant enzymes (54). This theory is consistent with the trapping of hydrolyzed myristic and lauric acids in alkyl channel A of the C143S mutant

enzyme, where they appear to be trapped (54). Interestingly, non-canonical binding of myristoyl-CoA is not observed in the OleA E117D structure, consistent with its ability to produce β-keto acid. The E117D OleA retains an acidic residue within alkyl channel B while E117A and E117Q enzymes do not. This suggests that the presence of a negative charge within alkyl channel B upon initial substrate binding may be important for proper positioning of the first substrate alkyl chain into the correct active site channel (alkyl channel A) for productive turnover. We also note that crystals of the mutants obtained in the absence of substrate are less well ordered than those of WT, as demonstrated by the observed poor diffraction, and consistent with the absence of Glu117 being associated with increased dynamics of the polypeptides.

Together these results show a clear reduction in product formation, a decrease in myristoyl-CoA hydrolysis, and a significant shift in the apparent pK_a when Glu117 is replaced by aspartate. Due to its conservation and position in the active site, we propose that Glu117 is the general base needed for proton abstraction to initiate condensation (Figure 28). When incubated with cerulenin, Cys143 from each mutant is able to be deprotonated and act as a nucleophile, properly orienting the alkylated product in alkyl channel A. This is consistent with the presence of an acidic residue not being absolutely required until after the acyl-enzyme intermediate is formed. The inability of OleA E117A and E117Q mutants to form condensed product suggests that Glu117 is directly involved during the Claisen condensation step (Figure 28). The active site residues of

OleA are equivalent to those of HMG-CoA synthase, which also employs a non-decarboxylative mechanism in which proton abstraction occurs using the general base, Glu79 (*Staphylococcus aureus* numbering) (23). However, Glu79 in HMG-CoA synthase is located on the same protein chain as the other active site residues, whereas Glu117 in OleA is located on the other polypeptide of the homodimer and is a novel feature for thiolase superfamily enzymes. The occurrence of the active site base originating from the other monomer causes the spatial orientation of the glutamate to shift in OleA relative to HMG-CoA synthase by ~5 Å (Figure 29A). Based on co-crystals of mutant OleA and myristoyl-CoA substrate, it has been proposed that Glu117 abstracts a proton from C_2 of the second non-acylated substrate (Figure 29B) (54). In contrast, Glu79 of HMG-CoA synthase has been proposed to abstract a proton from C_2 of the acetyl-Cys111 (Figure 29C) (23,53). This difference in substrate activation for condensation likely arises from the need for two alkyl coordinating channels, necessitating the shift in position of these bases and the glutamate not being sequentially equivalent between the two enzymes.

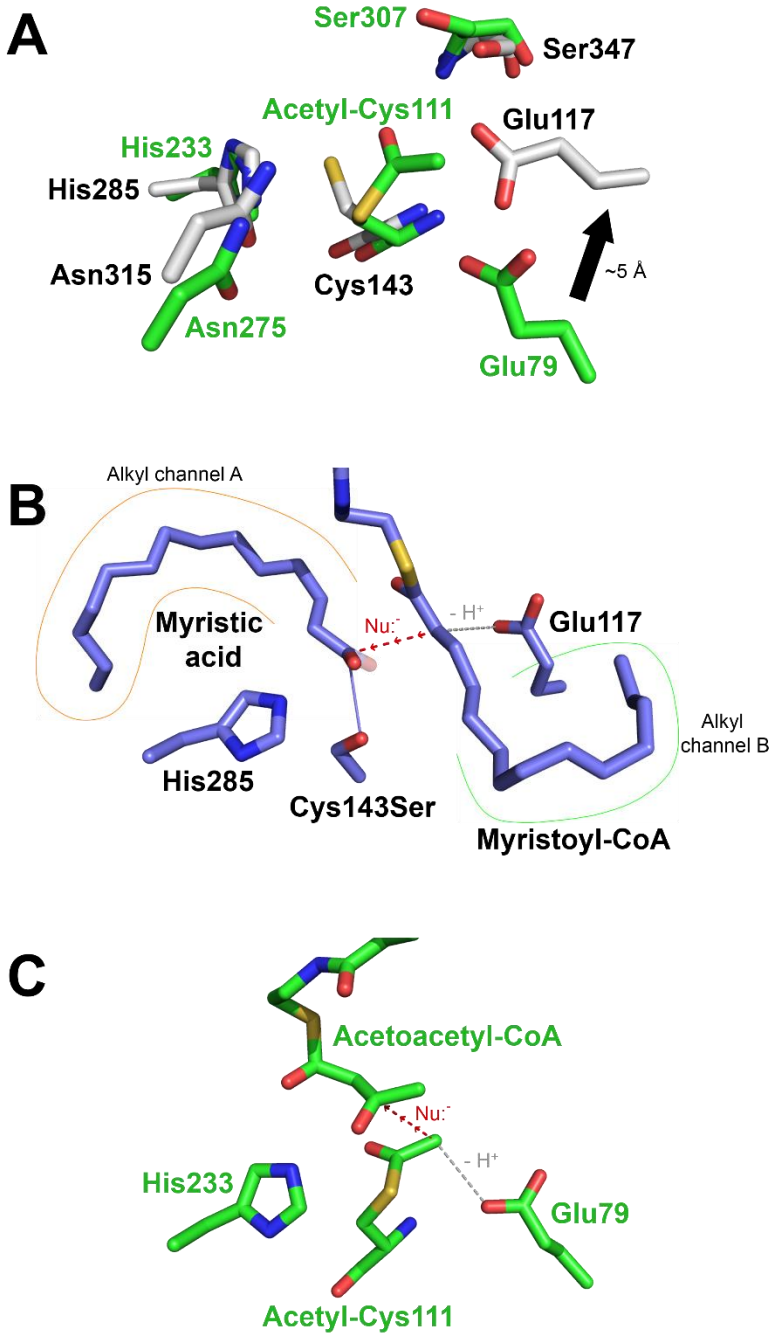


Figure 29: Active site comparisons between OleA and HMG-CoA synthase. **A)** Active site overlay between WT OleA (gray, PDB 3ROW) and HMG-CoA synthase (green, PDB 1XPK) *(23,25)*. HMG-CoA synthase is modelled with an acetylated Cys111. The active site glutamate base is shifted between the structures by an average of 5 Å due to the active site loop from which they originate. **B)** OleA C143S (purple, PDB 4KU3) active site substrate complex

showing substrate coordination. The purple solid line represents the position of the covalent bond between the first substrate (bound in alkyl channel A, orange) and the native Cys143 *(54)*. The gray dashed line represents the proposed deprotonation of the second substrate (bound in alkyl channel B, green) by Glu117. The red dashed line shows the likely path and direction of nucleophilic attack. **C)** HMG-CoA synthase active site poised for condensation between the acetylated Cys111 and acetoacetyl-CoA. The gray dashed line represents the priming deprotonation step. The red dashed line shows the path and direction of nucleophilic attack.

In the context of the thiolase superfamily, OleA appears to borrow features and characteristics from several of the functional subcategories. First, KAS and PKS type thiolases utilize a specificity pocket or hydrophobic channel to accommodate the growing acyl chain during condensation that is equivalent to alkyl channel B (21,33). In these enzymes the inhibitor cerulenin also binds in channel B (36). OleA utilizes not only this channel but a second alkyl channel as well for proper orientation of two long-chain acyl-CoA substrates, and cerulenin binds in alkyl channel A (54). Second, HMG-CoA synthases and biosynthetic thiolases prime condensation through a non-decarboxylative proton abstraction step. This is accomplished by an active site base, a glutamate for HMG-CoA synthase and a cysteine for biosynthetic thiolase (24,53). OleA appears to prime for condensation through deprotonating the nucleophilic substrate using Glu117. Third, we propose that the nucleophilic substrate for OleA is the second substrate that binds in alkyl channel B. KAS, PKS, and biosynthetic thiolases all utilize this strategy of nucleophilic attack, where the second substrate C_2 carbanion attacks the thioester of the acyl-enzyme intermediate (Figure 16A & C). Even though OleA and HMG-CoA synthase share a great deal of active site

residue orientation as well as both utilizing a glutamate for proton abstraction, structural arguments support OleA activating the second substrate for condensation rather than the acyl-enzyme intermediate as observed for HMG-CoA synthase.

In conclusion, we have solved the crystal structures for OleA E117A, E117Q, and E117D mutants and their complexes with cerulenin. Analyses of the myristoyl-CoA hydrolysis and product turnover rates, along with the apparent pK_a values for these reactions, show that these mutants have slowed myristoyl-CoA hydrolysis and are unable to form product if residue 117 is not an acid. Based on these data and spatial arguments from these and previous structures, we propose that Glu117 is the general base acting to deprotonate the second substrate, priming it for nucleophilic attack during the Claisen condensation step. Glu117 appears to also be important for orienting substrate within the appropriate alkyl binding channel for productive turnover. This is the first thiolase superfamily enzyme to utilize an active site base originating from the other monomer of the dimer. Compared to other thiolase enzymes, OleA seems to utilize mechanistic strategies from the three different subcategories. Thus, OleA represents a new hybrid subcategory of thiolase mechanisms.

CHAPTER 4

This chapter is has been submitted for publication in American Chemical Society *Biochemistry.*

The role of OleA His285 in coordination of long-chain acyl-coenzyme A substrates

Matthew R. Jensen, Brandon R. Goblirsch, Morgan A. Esler, James K. Christenson, Fatuma A. Mohamed, Lawrence P. Wackett, Carrie M. Wilmot

Chapter 4 summary

Renewable production of hydrocarbons is being pursued as a petroleum-independent source of commodity chemicals and replacement biofuels. The biosynthesis of long-chain olefins represents one platform for production of waxy hydrocarbons. The first enzyme in olefin biosynthesis, OleA, catalyzes the condensation of two fatty acyl-coenzyme A (CoA) substrates to form a β-keto acid. The mechanistic role of the absolutely conserved His285 was investigated through mutagenesis. H285A, H285N, and H285D OleA retain the ability to hydrolyze myristoyl-CoA, although the rate is severely compromised by the H285D mutation, but none are able to form condensed product. Complemented with crystal structures for each mutation and H285N OleA in complex with the thiolase inhibitor, cerulenin, these data suggest His285 interacts with Cys143 and the acyl-enzyme intermediate before and after transesterification. X-ray-induced

oxidation damage is observed in the His285 OleA variants suggesting His285 also has a protective role. Mechanistically, the residue is a key component of one of two oxyanion holes in the active site, so a role for substrate coordination is proposed in which oxyanion hole 1 is utilized by the first substrate followed by the second substrate using oxyanion hole 2. This mechanistic strategy is a reversal relative to thiolase enzymes such as 3-hydroxy-3-methylglutaryl-CoA synthase and FabH.

Introduction

The biosynthesis of long-chain hydrocarbons has been studied for decades, being initially described in 1929 as arising from the condensation of two fatty acids (66). These waxy compounds represent a renewable source of long-chain hydrocarbons for uses in replacement biofuels and commodity chemicals (4,48,49). The bacterial biosynthesis of long-chain olefins has become the subject of recent patents and investigations (7,9,10). The three- and four-gene clusters necessary for olefin production, *oleABCD*, have been identified in more than 250 divergent bacteria (Figure 30A) (7). The olefin biosynthesis pathway is initiated by the OleA-catalyzed condensation of two fatty acyl-coenzyme A (CoA) molecules to form a β-keto acid (Figure 30B) (15). The subsequent NADPH-dependent reduction of β-keto acid to a β-hydroxy acid is catalyzed by OleD (16). The enzyme OleC acts as a β-lactone synthetase, catalyzing an ATP-dependent reaction of the β-hydroxy acid to form a β-lactone (50). The final cis-olefin is

produced by OleB through displacement of carbon dioxide from the β-lactone (20).

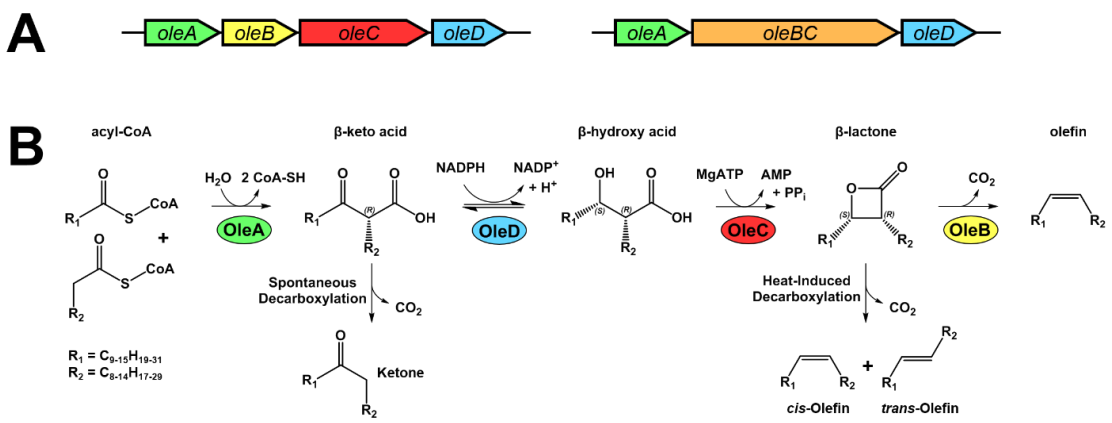


Figure 30: Proposed mechanism of bacterial olefin biosynthesis. **A)** Representative four-gene (left) and three-gene (right) cluster arrangements for *oleABCD*. The cluster containing an *oleBC* fusion (orange) is found in olefin-producing actinobacteria. **B)** Mechanistic steps of olefin production catalyzed by OleA, OleD, OleC, and OleB. Typical acyl-CoA substrates are between 10 and 16 carbons in length in *Xanthomonas campestris*. Figure produced from (7,15,16,20,50).

The mechanism of OleA has been the subject of several studies aiming to understand which amino acid residues and structural features of the enzyme are important for function. OleA demonstrate substrate specificity toward different acyl-CoA molecules that ultimately determine the chain length, unsaturation, or other modifications of the alkyl chain moieties present in the final olefin (7). As such, OleA represents an important target for enzyme engineering efforts for tailored hydrocarbon production. OleA is a member of the thiolase superfamily. It condenses two acyl-CoA substrates in a non-decarboxylative reaction to generate a β-keto acid and two CoA-SH (15). OleA from *Xanthomonas*

campestris has been characterized structurally, showing a novel T-shaped, three-channel substrate binding architecture centered around the catalytic cysteine (Cys143) that is necessary for positioning the hydrophobic substrate alkyl chains for condensation (25,54). Another novel feature of OleA among thiolase enzymes is the presence of an active site glutamate (Glu117) originating from the second monomer of the OleA homodimer (25). Mutagenesis of this glutamate, paired with crystal structures of inhibitor complexes and substrate modelling, suggest that Glu117 primes the condensation reaction by deprotonating the second substrate for nucleophilic attack (Chapter 3). The current proposed mechanism for OleA involves three steps (Figure 31) (Chapter 3). The transesterification step proceeds via a nucleophilic attack on the first acyl-CoA substrate by the Cys143 thiolate, forming an acyl-enzyme intermediate (Figure 31A). This intermediate occupies alkyl channel A during turnover (54). Claisen condensation occurs after Glu117 abstracts a proton from C_2 of the second substrate, which is bound in alkyl channel B (Figure 31). This carbanion species attacks the thioester linkage of the acyl-enzyme intermediate, forms a new carbon-carbon bond, and releases the intermediate from Cys143. The final hydrolysis of the β-keto acyl-CoA then occurs before product release, likely catalyzed by an activated water molecule (Figure 31C) (Chapter 3).

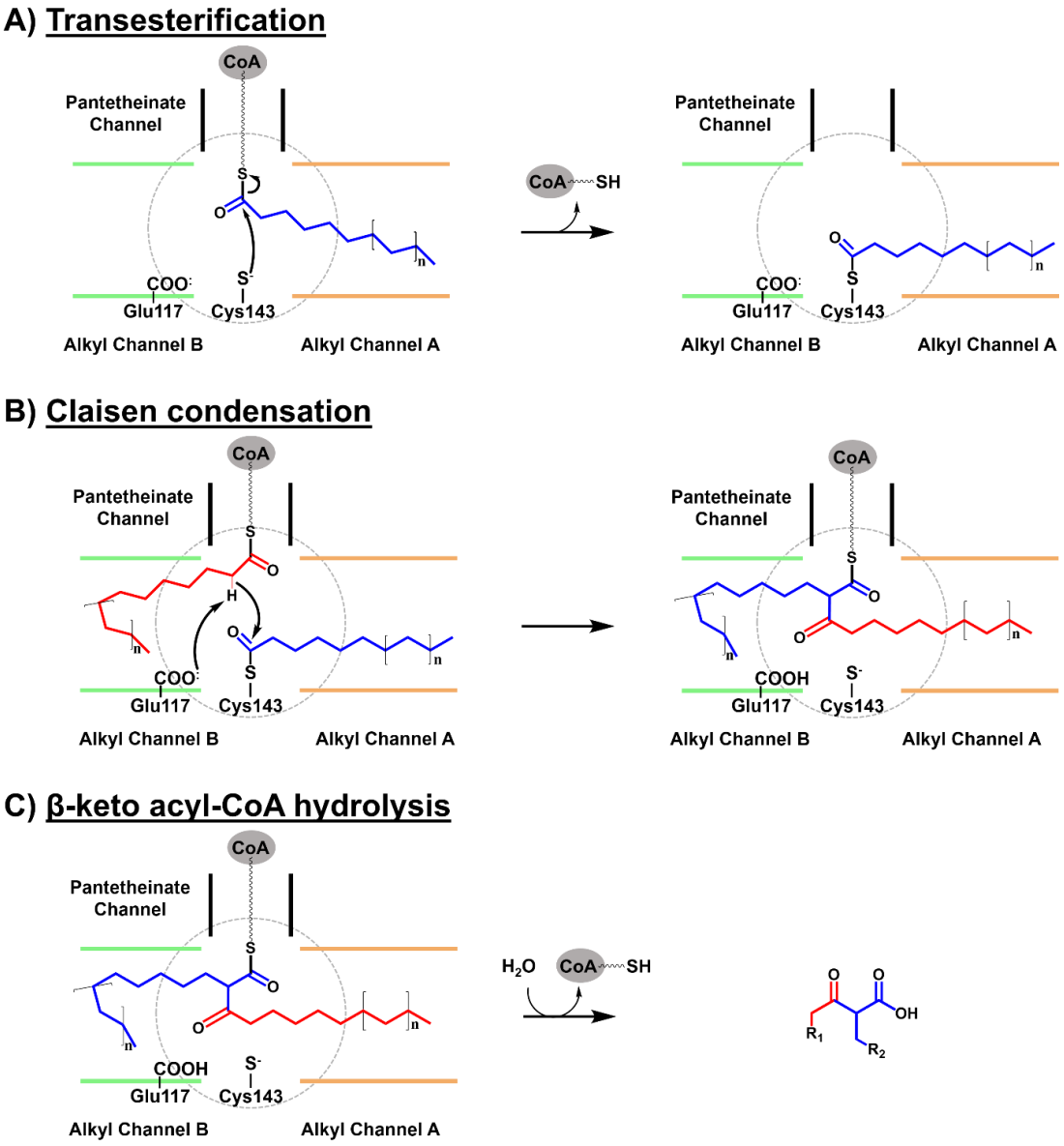


Figure 31: Proposed mechanism of OleA. **A)** Transesterification: the first acyl-CoA (blue) binds in alkyl channel A (orange) forming an acyl-enzyme intermediate after nucleophilic attack by Cys143. **B)** Claisen condensation: Glu117 abstracts a proton from C_2 of the second substrate (red) bound in alkyl channel B (green), priming nucleophilic attack between the carbanion and the acyl-enzyme intermediate. **C)** β-keto acyl-CoA hydrolysis: the final CoA-SH is released through hydrolytic displacement by a hypothesized activated water.

All thiolase enzymes contain four active site loops that are necessary for aligning catalytic residues. The first loop contains the catalytic cysteine, and the second and third loops contain important residues such as histidine and asparagine (14). A main chain amide group on the fourth loop forms an oxyanion hole that stabilizes the negatively charged intermediate during the conserved acetyl transfer/transacylation/transesterification step (14). In some superfamily members, the side chains of the residues found on loops two and three (typically histidine and/or asparagine) form a second oxyanion hole that binds the second substrate and stabilizes subsequent intermediates during condensation. A histidine is found within the active site of all thiolase enzymes with a role proposed in most mechanisms. In β-ketoacyl-acyl-carrier protein synthases (KAS; FabH, FabB) and polyketide synthases (PKS; chalcone synthase), an active site histidine is proposed to promote decarboxylation of the second substrate, malonyl-acyl carrier protein (ACP) (67,68). In 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase), the histidine is proposed to act as a general base to deprotonate the catalytic cysteine for nucleophilic attack on the first substrate, acetyl-CoA (28). This same histidine is also implicated in selectively interacting with the second substrate, acetoacetyl-CoA (69). In biosynthetic thiolase, the histidine is proposed to act as a general base to deprotonate the catalytic cysteine, as well as a general acid to donate a proton back to the cysteine to complete turnover (24). For OleA, the mechanistic role of this histidine (His285 in *X. campestris*) is unknown. In the wild type (WT) crystal

structure of OleA, the side chain of His 285 forms part of an oxyanion hole (oxyanion hole 1) alongside the side chain of Asn315 (25).

In this study, we have investigated the role of His285 in *X. campestris OleA* through site-directed mutagenesis to generate H285A, H285N, and H285D OleA variants. Due to the residue being absolutely conserved across OleA enzymes as well as its spatial proximity to Cys143, His285 was hypothesized to play a role in activating transesterification (25). Crystal structures of these mutant enzymes and the cerulenin-inhibited H285N variant were determined. Structural and biochemical analyses of these mutants suggest roles for His285 in catalytic intermediate stabilization and positioning, as well as protection from oxidative damage. These results are placed in the context of the OleA mechanism and those of other thiolases.

Experimental

Mutagenesis, expression, and purification of OleA His285 site mutation

Site-directed mutagenesis of *E. coli* codon-optimized *Xanthomonas campestris* pv. *campestris* str. ATCC 33913 (NP_635607.1) *oleA* gene was carried out as previously described to introduce the H285N site mutation (54). H285A and H285D *oleA* mutations were constructed from *E. coli* codon-optimized gBlocks purchased from Integrated DNA Technologies (IDT, Coralville, IA). Synthetic DNA was integrated into pET28b^{+} expression vectors using Gibson

assembly (70). Mutations and vector constructions were verified by DNA sequencing (ACGT, Inc.).

His285 mutated OleA proteins were expressed in BL21 (DE3) *Escherichia coli* competent cells (New England Biolabs). Cells were grown at 37°C in flasks containing 1 L LB media and 50 µg/mL kanamycin sulfate until an A600 between 0.5-0.7 was achieved. Expression was induced by addition of 100 µM isopropyl β-D-1-thiogalactopyranoside. After 4 hr of expression at 37°C, cells were harvested by centrifugation at 4000 x *g* for 10 min.

Following resuspension in buffer consisting of 200 mM sodium chloride and 20 mM Tris HCl pH 7.4, cells were lysed by sonication (QSonica Q700). Soluble lysate was isolated by centrifugation at 27,000 x *g* for 1 hr and filtered through a 0.2 µm syringe filter. Purification of the His_6-tagged variants was carried out on an ÄKTA FPLC (GE Healthcare Life Sciences) using nickel-affinity chromatography (5 mL HisTrap HP column) and size exclusion chromatography (HiLoad 16/600 Superdex 200 pg column). The resulting enzymes were buffer exchanged into 200 mM sodium chloride and 20 mM Tris HCl pH 7.4. The purity of His285 OleA variants were analyzed by SDS-PAGE using SimplyBlue SafeStain (Invitrogen). Mutations were confirmed by mass spectrometry using a Bruker Autoflex Speed Matrix-assisted laser desorption/ionization time of flight mass spectrometer (MALDI-TOF) (University of Minnesota Center for Mass Spectrometry and Proteomics). Purified protein was concentrated to 115 µM (9 mg/mL) for crystallization trials.

OleA mutant enzyme hydrolysis of acyl-CoA

Detection of free CoA-SH hydrolyzed from myristoyl-CoA by OleA was determined as previously described (54). Briefly, WT and His285 OleA variants (20 µg) were incubated with 65 µM myristoyl-CoA for times ranging from 15 min to 24 hr in 500 µL of buffer consisting of 200 mM NaCl, 4% v/v ethanol, and 20 mM Tris HCl pH 7.4. Reactions were quenched by an equal volume addition of 8.25 M guanidinium HCl at pH 7.4 before a 10 min incubation with 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB, 300 µM). Free thiol resulting from OleA-dependent hydrolysis was quantified spectrophotometrically at 412 nm (ε_{412}=13,600 $M^{-1}cm^{-1}$) (15,34,35).

Activity assay for condensation of acyl-CoA

Detection of the OleA condensation product, β-keto acid, was carried out using gas chromatography-mass spectrometry (GC/MS). Under the high heat conditions of the GC inlet, β-keto acids thermally decarboxylate quantitatively to the corresponding ketone (15). OleA WT and His285 variant proteins (10 µg) were reacted at room temperature in glass vials containing 50 µM myristoyl-CoA and buffer consisting of 200 mM NaCl, 4% v/v ethanol, and 20 mM Tris HCl pH 7.4 (500 µL total volume). Following incubations of 15 min to 24 hr, product was extracted with 500 µL ethyl acetate. The organic phase was analyzed by GC/MS (5 µL injection split between flame ionization detector (FID) and mass spectrometer) using the following parameters: 250°C inlet temperature, 100°C start temperature, 10°C/min ramp rate, 320°C final temperature, and a 5 min hold

time (total time of 27 min). Product elution time and mass spectral fingerprints were confirmed using the commercially available ketone standard, 14-heptacosanone.

Crystallization

Crystallization conditions for WT OleA were described previously (25). All crystallizations of H285A, H285N, and H285D OleA were set up using hanging drop vapor diffusion, combining 1 µL of protein with 1 µL of well solution. Each variant protein was crystallized alone, co-crystallized with myristoyl-CoA (1 mM), and co-crystallized with cerulenin (1 mM). Protein-alone and myristoyl-CoA co-crystal well solution conditions were as follows: 0.1 M sodium citrate pH 4.2, 0.10-0.12 M potassium phosphate dibasic, and 14-20% PEG 8000. Cerulenin co-crystal well solution conditions were as follows: 0.07-0.10 M manganese (II) chloride, 0.1 M MES pH 6.0, and 13-17% w/v PEG 4000. Crystal trays were incubated at 19°C for 1-3 days before crystals appeared. Crystals were harvested, cryo-protected in well solution containing 25% v/v glycerol, and flash-frozen in liquid nitrogen.

X-ray data collection, processing, and refinement

H285A, H285N, and H285D OleA datasets were collected at GM/CA-CAT beamline 23-ID-D (Pilatus3 6M) and at SBC-CAT beamline 19-BM (ADSC Quantum 210r) at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Crystals were exposed to 0.97, 1.02, or 1.03 Å wavelength X-rays with an attenuated beam (1-5 fold) at 100 K. Datasets were processed and

scaled using either XDS or HKL2000 (58,59). The WT OleA crystal structure (PDB 3ROW, 1.85 Å) was used to phase the isomorphous H285A, H285N, and H285D myristoyl-CoA co-crystal datasets by direct Fourier synthesis (CCP4 Suite, REFMAC5) (60,71). The H285N OleA-cerulenin co-crystal dataset was phased by molecular replacement (PhaserMR) using the WT OleA-cerulenin crystal structure (PDB 3S21, 1.7 Å) as an initial search model (62). Equivalent reflections to those used for R_{free} determination in the WT and WT-cerulenin structure determinations were used in myristyol-CoA and cerulenin co-crystal refinements, respectively (25). Model building was carried out in COOT with refinement after each round using REFMAC5 until all interpretable electron density was explained (63). Adjustments to residue occupancies for alternative conformers were made in 5% increments until B factor values were similar between conformers and surrounding amino acids post-refinement. A final round of restrained refinement with TLS was carried out before structural validation using MolProbity (46). Cerulenin ligand atoms outside of interpretable electron density were manually changed to zero occupancy following the final round of refinement. Final structures were deposited in the Protein Data Bank (PDB), and all figures depicting protein structures were prepared in PyMOL (47). Superimpositions and root mean square deviations (rmsd) of crystal structures were calculated using PyMOL using the SUPER command over all α-carbons.

Results

Mutagenesis, expression, and purification of OleA H285 variants

The residue His285 is absolutely conserved in OleA primary amino acid sequences (Figure 32). Its role in the formation of β-keto acid was investigated through mutation of His285 and subsequent characterization of the resulting OleA variants. H285A, H285N, and H285D mutations of OleA were selected to remove side chain interactions, maintain hydrogen bonding capabilities but remove ionization, and introduce a negatively charged amino acid side chain, respectively. All variant and WT OleA proteins were expressed and purified to >95% purity as estimated by SDS-PAGE. H285A and H285N OleA expressions yielded >20 mg/L of cell culture while H285D OleA yielded ~15 mg/L of cell culture. The mass of each enzyme variant was confirmed using MALDI-TOF mass spectrometry.

X. campestris	D	Q	F	V	I	^{285}**H**	Q	V	S	R	P
G. uraniireducens	D	R	V	I	C	^{274}**H**	Q	V	G	S	A
S. oneidensis	D	K	V	I	C	^{276}**H**	Q	V	G	A	S
S. ambofaciens	S	F	I	A	F	^{269}**H**	Q	I	A	M	P
A. aurescens	D	R	Y	V	T	^{269}**H**	Q	V	S	N	A
K. radiotolerans	E	K	F	V	I	^{285}**H**	Q	I	S	K	V
M. luteus	D	R	F	I	T	^{272}**H**	Q	V	S	Q	M
S. maltophilia	D	Q	F	V	I	^{265}**H**	Q	V	S	Q	P
C. aggregans	A	L	Y	A	P	^{266}**H**	Q	V	G	A	R

Figure 32: Sequence alignment of select OleA enzymes from divergent bacteria highlighting the invariant histidine residue. *Xanthomonas campestris* pv. *campestris* str. ATCC 33913, *Geobacter uraniireducens* Rf4, *Shewanella oneidensis* MR-1, *Streptomyces ambofaciens* ATCC 23877, *Arthrobacter aurescens* TC1, *Kineococcus radiotolerans* SRS30216, *Micrococcus luteus* NCTC 2665, *Stenotrophomonas maltophilia* K279a, *Chloroflexus aggregans* DSM 9485.

Effect of mutation of His285 on OleA hydrolysis of acyl-CoA

The concentration of free CoA-SH produced by the OleA-dependent futile or productive hydrolysis of myristoyl-CoA was detected spectrophotometrically using DTNB. As previously reported, WT OleA completely hydrolyzes all available substrate in under 15 min (Table 9) (Chapter 4). H285A and H285N OleA both hydrolyze 98% and 100% of myristoyl-CoA in the same amount of time as WT OleA (Table 9). Only the hydrolysis rate for the H285D mutation is affected, needing 24 hr to produce similar levels of CoA-SH as WT, H285A, and H285N OleA.

Table 9: Hydrolysis of myristoyl-CoA by WT OleA and His285 variants.
Values shown are the average in triplicate with standard deviation.

Enzyme	Enzyme concentration (μM)	Rxn time (hrs)	CoA product (μM)[a]	Percent yield (%)[b]
WT	0.5	0.25	32 ± 0.1	100
H285A	0.5	0.25	31 ± 0.2	98
H285N	0.5	0.25	33 ± 1.3	100
H285D	0.5	0.25	1.1 ± 0.2	3
H285D	0.5	24	29 ± 0.1	91

[a] Free coenzyme A detected as described under "EXPERIMENTAL"
[b] Starting substrate was 32.5 μM; 32.5 μM product is 100% turnover

Effect of His285 mutation on OleA condensation of acyl-CoA

The thermally labile β-keto acid product is detected as a stable ketone (14-heptacosanone) by GC/MS. The product of WT OleA reacted with myristoyl-CoA shows a product that elutes at 16 min, identical in retention time and mass spectrum to 14-heptacosanone (Chapter 3). After 15 min incubation with substrate, WT OleA consumes nearly 90% of available substrate (Table 10). All

His285 OleA variants are unable to condense product. After a 24 hr incubation, H285A, H285N, and H285D OleA produce no discernable elution peak at 16 min (Table 10).

Table 10: Production of β-keto acid by WT OleA and His285 variants.

Enzyme	Enzyme concentration (μM)	Rxn time (hrs)	Ketone peak area (a.u.)[a]	Normalized ketone peak area (%)[b]
WT	0.25	0.25	5599 ± 146	86
WT	0.25	24	6510 ± 156	100
H285A	0.25	24	N.D.[c]	N.D.
H285N	0.25	24	N.D.	N.D.
H285D	0.25	24	N.D.	N.D.

[a]Values shown are the average in triplicate of the GC-FID peak area (arbitrary units) with standard deviation.
[b]GC-FID peak area as a percentage of WT 24 hour reaction peak area.
[c]N.D.; not detectable: peak area below 50 a.u. considered indistinguishable from baseline.

Effect of His285 mutations on OleA structure

Statistics for data collection and refinement of each crystal structure are shown in Table 11. The crystal structures of H285A, H285N, and H285D OleA were solved to 1.77 Å, 2.0 Å, and 2.7 Å resolution, respectively. Each was solved in the $P2_12_12_1$ space group with the physiological dimer present as the crystallographic asymmetric unit. The overall structures of H285A, H285N, and H285D OleA show little perturbation compared to the WT structure (PDB 3ROW) with an rmsd across all α-carbon atoms of 0.146 Å, 0.201 Å, and 0.202 Å, respectively. Although co-crystallized in the presence of myristoyl-CoA, there was no evidence of bound substrate, product, or chemical derivative in the electron density. Crystals grown in the absence of substrate produced inferior diffraction data, suggesting increased disorder. The structure of H285N OleA in complex with cerulenin was solved to 2.04 Å and exhibits an rmsd of 0.271 Å to

WT OleA-cerulenin (PDB 3S21). As with WT OleA-cerulenin, H285N OleA-ceulenin was solved in the $P3_221$ space group with a single monomer in the asymmetric unit. Crystals of H285A OleA-cerulenin and H285D OleA-cerulenin were produced; however, the crystals diffracted poorly.

Table 11: Data collection and refinement statistics for OleA His285 variants.

	H285A OleA	H285N OleA	H285D OleA	H285N OleA–Cerulenin
Data Collection				
Wavelength (Å)	1.02	1.03	0.97	1.03
Space group	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$	$P\,2_1\,2_1\,2_1$	$P\,3_2\,2\,1$
Unit cell (Å)	82.2 x 85.6 x 103.0	81.4 x 84.7 x 103.2	81.4 x 85.0 x 103.0	90.2 x 90.2 x 69.1
Unit cell (°)	α = β = γ = 90	α = β = γ = 90	α = β = γ = 90	α = β = 90; γ = 120
Resolution (Å)	50.0-1.77 (1.80-1.77)	29.72-2.00 (2.06-2.00)	50.0-2.70 (2.77-2.70)	29.51-2.04 (2.09-2.04)
Measured reflections	488250	208959	132671	203550
Unique reflections	71384	48443	20266	20876
Completeness (%)	99.5 (96.9)	98.9 (92.6)	99.6 (95.0)	99.0 (95.5)
R_{merge} (%)[b]	9.4 (87.7)	6.2 (45.5)	14.4 (54.6)	5.8 (72.8)
I/σ_I	14.6 (2.6)	16.0 (2.8)	10.5 (2.4)	24.2 (2.9)
CC*	0.976 (0.903)	0.999 (0.947)	0.998 (0.974)	0.999 (0.926)
Multiplicity	6.8 (5.6)	4.3 (3.8)	6.5 (6.7)	9.8 (9.0)
Wilson B factor (Å^2)	21.0	29.6	40.6	36.3
Source	APS 19 BM	APS 23 ID-D	APS 23 ID-D	APS 23 ID-D
Processing software	HKL2000	XDS	HKL2000	XDS
Refinement				
Resolution (Å)	1.77	2.00	2.70	2.04
Monomers in ASU	2	2	2	1
R_{work}[c]	0.1749	0.1641	0.1722	0.1786
R_{free}[d]	0.2091	0.2182	0.2469	0.2300
Ramachandran statistics (%)[e]				
favored	97.2	96.4	95.6	96.0
outliers	0.15	0.59	0.30	0.00
RMS deviation				
bond lengths (Å)	0.0259	0.0242	0.0113	0.0105
bond angles (°)	2.373	2.258	1.555	1.288
Average B-factor (Å^2)	23.0	31.7	41.7	39.8
ligand B-factor (Å^2)	–	–	–	95.5
Clashscore (percentile)[e]	1.78 (100th)	2.26 (99th)	2.46 (100th)	4.09 (99th)
Protein Data Bank ID	Not yet deposited	Not yet deposited	Not yet deposited	Not yet deposited

[a]Data in parentheses are for the highest resolution shell. [b]$R_{merge} = \Sigma_i\, |I_{hkl,\,i} - \{I_{hkl}\}| / \Sigma_{hkl}\, \Sigma_i\, I_{hkl,\,i}$, where **I** is the observed intensity and $\{I_{hkl}\}$ is the average intensity of multiple measurements. [c]$R_{work} = \Sigma||F_o|-|F_c|| \,/\, \Sigma|F_o|$, where $|F_o|$ is the observed structure factor amplitude and $|F_c|$ is the calculated structure factor amplitude. [d]R_{free} is the R factor based on 5% of the data excluded from refinement. [e]Based on values obtained from MolProbity (46).

Effect of His285 mutation on overall OleA structures

The structures of H285A, H285N, and H285D OleA show little perturbation to the three-channel architecture compared to WT. In all three crystal structures, alternative conformers and oxidation of Cys143 are observed in at least one monomer active site (Figure 33). Simulated annealing omit maps were utilized during modelling, and were consistent with oxidative modification of the active site cysteine (Figure 33B). MALDI-TOF MS of H285A, H285N, and H285D OleA in solution showed the expected mass with no evidence of oxidation (data not shown). The crystal structure of H285N OleA purified and crystallized in the presence of 2-mercaptoethanol similarly shows oxidation of Cys143 (data not shown). These data suggest that cysteine oxidation occurs during X-ray data collection. In monomer A of H285A OleA, alternative conformers of Cys143 are present at 80% and 20% occupancy differing by ~112° (Figure 33C). In monomer B, one conformer of Cys143 is oxidized to sulfenic acid (45% occupancy) while the other remains unmodified (55% occupancy) (Figure 33D). The water molecule bound in oxyanion hole 2 – formed by the backbone amides of Cys143 and Ser347 – is bound at 55% occupancy and is likely displaced by the hydroxyl of the oxidized conformer of Cys143. A water molecule found in both monomer active sites of H285A OleA forms a hydrogen bond with the side chain carbonyl of Asn315 (Figure 33C & D). Compared to WT OleA, this water occupies the position of the native His285 imidazole ring (Figure 33A, C, & D). H285N OleA shows oxidation in both monomers (Figure 33E & F). Monomer A contains a

singly-oxidized Cys143 with the hydroxyl group bound in oxyanion hole 2 (Figure 33E). Two conformers of sulfenic acid-Cys143 are present in monomer B at 50% occupancy (Figure 33F). The hydroxyl of one conformer is bound in oxyanion hole 2 while the other points into the pantetheinate channel. The side chain of Asn285 forms two hydrogen bonds with the main chain amide and carbonyl of Ile345 at the base of alkyl channel A (Figure 34). Similar to H285A OleA, H285D OleA shows oxidation of Cys143 exclusively in monomer B (Figure 33G & H). Monomer A of the H285D mutant has two conformers of Cys143 at 60% and 40% occupancy with a water molecule trapped in oxyanion hole 2 (Figure 33G). Two conformers of oxidized Cys143 are present in H285D monomer B at similar occupancies (Figure 33H). Due to a hydrogen bond to one sulfenic acid hydroxyl, the water molecule bound in oxyanion hole 2 loses the canonical hydrogen bond to Ser347 observed in WT, H285A, and H285N OleA. The side chain of Asp285 also forms one hydrogen bond to the main chain amide of Ile345 (not shown).

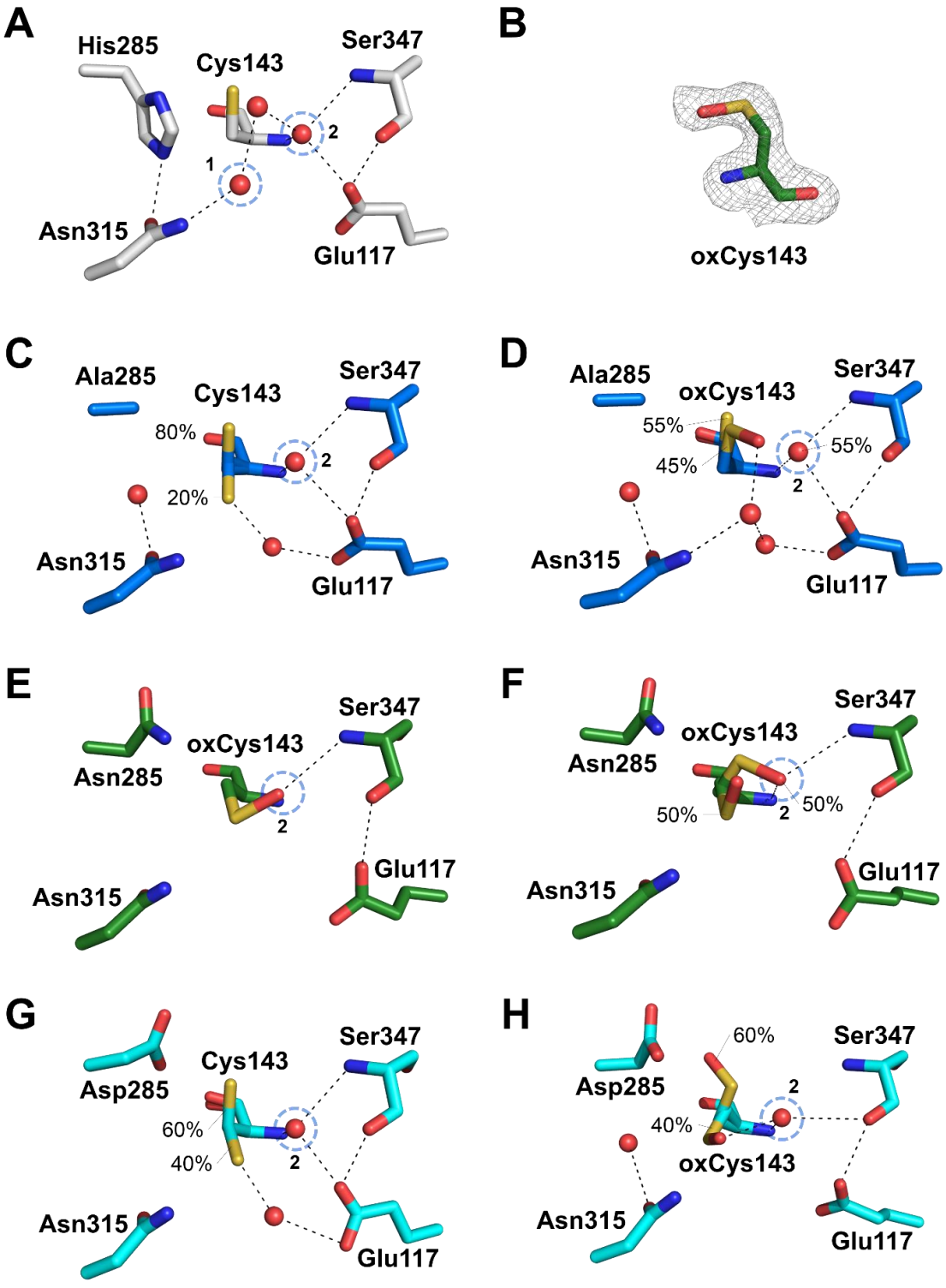


Figure 33: Active site comparisons of OleA His285 variant structures. **A)** WT OleA active site residues (gray) showing water molecules (red spheres) bound in

oxyanion holes 1 and 2 (blue dashed circles) (25). **B)** Example simulated annealing omit $mF_o - DF_c$ electron density (contoured at 3σ) from H285N OleA structure (green) used to assign oxidation of Cys143. **C)** Monomer A active site of H285A OleA (blue). The two conformers of Cys143 are modelled at 80% and 20%. **D)** Monomer B active site of H285A OleA (blue). Cys143 is modelled as 55% cysteine and 45% sulfenic acid. The water molecule bound in oxyanion hole 2 is modeled at 55% occupancy to match that of the unmodified cysteine. **E)** Monomer A active site of H285N OleA (green). Cys143 is oxidized to sulfenic acid with the hydroxyl group bound in oxyanion hole 2. **F)** Monomer B of H285N OleA (green). Cys143 is modelled as two conformers of sulfenic acid each at 50% occupancy. **G)** Monomer A active site of H285D OleA (cyan). The two conformers of Cys143 are modelled at 60% and 40%. **H)** Monomer B of H285D OleA (cyan). Cys143 is modelled as 60% and 40% sulfenic acid conformers. All hydrogen bonds are depicted as black dashed lines.

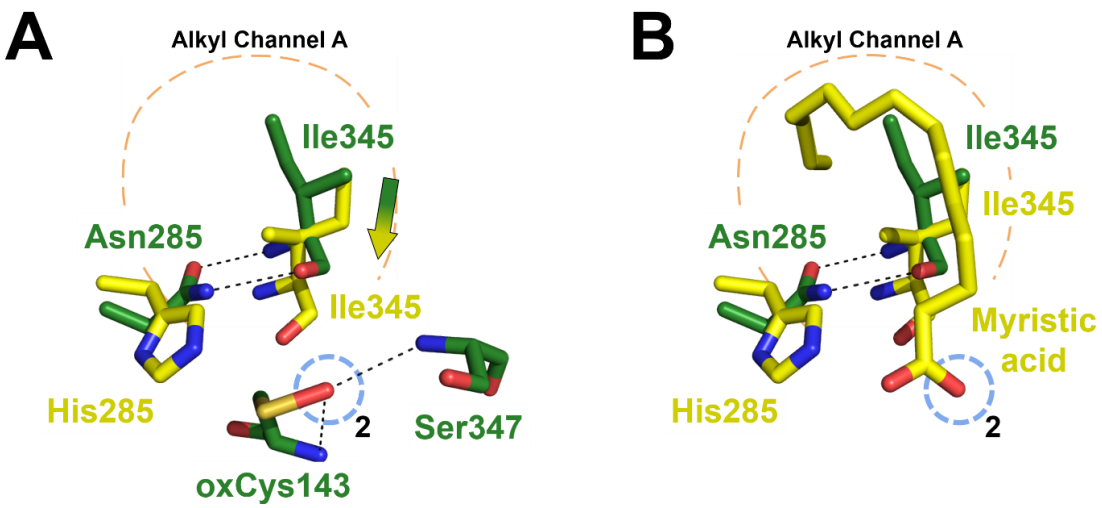


Figure 34: Asn285 side chain interactions near alkyl channel A. **A)** Overlay of H285N OleA (green) and C143S OleA (yellow, PDB 4KU3) (54). Asn285 (green) forms two hydrogen bonds with the main chain of Ile345 that do not allow it to shift to the substrate-bound conformation that Ile345 (yellow) adopts when alkyl channel A is occupied. **B)** Overlay of H285N OleA (green) and C143S OleA (yellow, PDB 4KU3) showing the conformational changes around Ile345 needed to position myristic acid (yellow) within alkyl channel A (54).

Crystal structure of H285N OleA bound to the thiolase inhibitor cerulenin

Cerulenin is a covalent inhibitor of thiolases that reacts with the active site cysteine nucleophile. Unlike the WT OleA-cerulenin complex, the H285N OleA -

cerulenin structure does not contain a covalent bond between Cys143 and cerulenin (Figure 35A & B). In the crystal structure, cerulenin refines to a distance that is too far for a covalent bond between the cerulenin C_2 atom and the thiol group of Cys143 (3 Å) (Figure 35C). It is likely that the carbon-sulfur bond is broken by radiation damage during X-ray data collection, as MALDI-TOF MS analysis indicates cerulenin is covalently bound to H285N OleA in solution (data not shown). Strong, positive $mF_o - DF_c$ electron density and simulated annealing omit maps suggest that cerulenin is still bound in the active site where it maintains similar hydrogen bonding interactions as those observed in the WT OleA-cerulenin structure (Figure 35A). The carboxamide of cerulenin makes two hydrogen bonds with Ser347 and a third with the main chain amide of Cys143 (Figure 35B). Unlike WT OleA-cerulenin, the C_4 carbonyl oxygen of cerulenin forms a hydrogen bond with Asn315, which forms oxyanion hole 1 with the side chain of His285. In WT OleA-cerulenin, this hydrogen bond is formed between Asn315 and the C_3 hydroxyl group (Figure 35A). In H285N OleA-cerulenin, Asn285 does not form the second half of oxyanion hole 1 as it is rotated 60° relative to His285 in WT OleA-cerulenin (Figure 35A & B). The alkyl chain of cerulenin exhibits a great deal of flexibility within alkyl channel A based on the lack of interpretable electron density, likely primarily due to the loss of the

covalent linkage (Figure 35B & C).

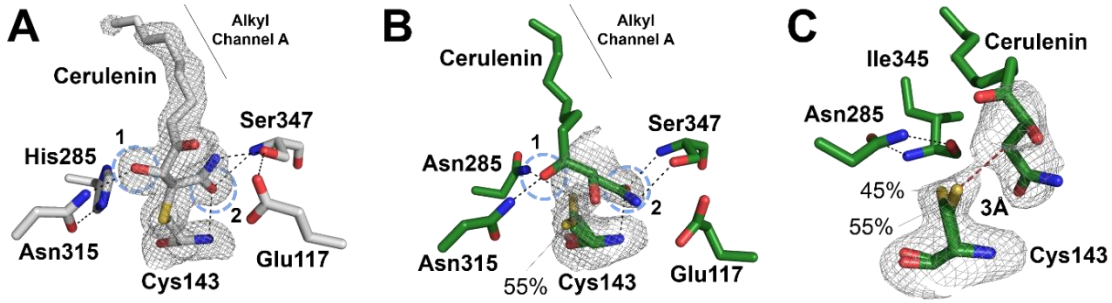


Figure 35: Comparison of active site residues of WT and H285N OleA bound with cerulenin. **A)** WT OleA Cys143 bound covalently with cerulenin. Hydrogen bonds are shown as black dashed lines. The alkyl chain of cerulenin is bound in alkyl channel A (25). **B)** H285N OleA active site with cerulenin bound in alkyl channel A. Cys 143 has two conformers at 55 and 45% occupancy. Cerulenin shares the same hydrogen bonds as WT OleA-cerulenin except the ligand is not covalently bound to Cys143. **C)** View of cerulenin as seen looking into alkyl channel A. C_2 of cerulenin in 3 Å from the thiol of Cys143 (red dashed line), indicating that it is not covalently bound to the enzyme. Asn285 makes two hydrogen bonds to the main chain of Ile345. Simulated annealing omit maps are contoured at 3σ.

Discussion

We report the crystal structures of H285A, H285N, H285D OleA, and H285N OleA-cerulenin. Oxidation of Cys143 to sulfenic acid is observed in each unbound OleA His285 variant structure, which is the first example of oxidative damage to Cys143 observed in OleA (25,54) (Chapter 3). The observed oxidation likely occurs during X-ray diffraction data collection. MALDI-TOF MS showed the non-oxidized mass of each variant, and the H285A and H285N OleA variants are able to hydrolyze myristoyl-CoA at the same rate as WT. The oxidation observed in the His285 variant crystal structures indicates that His285 has a protective role through its interaction with Cys143. Monomer B of the

His285 variants is surprisingly more susceptible to this oxidative damage, possibly due to fewer crystal contacts that allow for increased disorder compared to monomer A (Chapter 3). The H285N OleA-cerulenin structure does not contain oxidation of Cys143 but rather shows radiative damage to the trapped ligand. In solution, H285N OleA-cerulenin shows a similar mass shift to WT OleA-cerulenin consistent with covalent modification of the enzyme. In the crystal structure, the carbon-sulfur bond appears to be broken as C_2 of cerulenin refines 3 Å from the thiolate of Cys143. It is well documented that carbon-sulfur bonds are among the first to delocalize when protein crystal structures at 100 K are irradiated (72-74). However, the absence of His285 makes this bond more susceptible to photoreduction as it remains intact in WT and Glu117 variant crystal structures, consistent with His285 having a protective role in mitigating redox damage (25) (Chapter 3). Cerulenin is still bound non-covalently in the active site where it hydrogen bonds to several active site residues (Figure 35B). However, unlike His285 in OleA Glu117 variant structures, Asn285 does not form a hydrogen bond directly with cerulenin due to the formation of two hydrogen bonds with Ile345 (Figure 35C) (Chapter 3). Compared to WT, the increased disorder observed in the alkyl chain of cerulenin is likely a result of the loss of covalent attachment to Cys143 as the amino acids of alkyl channel A are structurally equivalent to those observed in the WT-cerulenin structure.

The role of His285 in the mechanism of OleA was investigated due to its conservation and the importance of histidine residues in other thiolases. It was

initially hypothesized that His285 acts as a general base to deprotonate Cys143 for the transesterification reaction (25). A similar proton abstraction role has been proposed in KAS type I and II mechanisms where the histidine is within favorable hydrogen bonding distance to the catalytic cysteine thiol (37,75,76). In WT OleA, the imidazole ring of His285 is too far from the thiol of Cys143 (3.5 Å) to form a strong hydrogen bond, and it is at an unfavorable angle to abstract a proton (Figure 36A) (25). It is likely that Cys143 is activated for transesterification through its position directly adjacent to the N-terminal dipole moment of α-helix 4 (Figure 36). The effect this dipole exerts on the electrostatic environment of condensing enzymes have been implicated in many thiolase superfamily members (77-79). The magnitude of this dipole is suggested to lower the pK_a of the catalytic cysteine sufficiently to allow for deprotonation. However, His285 may stabilize the formation of an imidazolium-thiolate ion pair as proposed in chalcone synthase and cysteine proteases (68,80). This interaction could protect Cys143 from becoming oxidized prior to transesterification, and is consistent with the susceptibility of Cys143 to oxidation in the absence of His285.

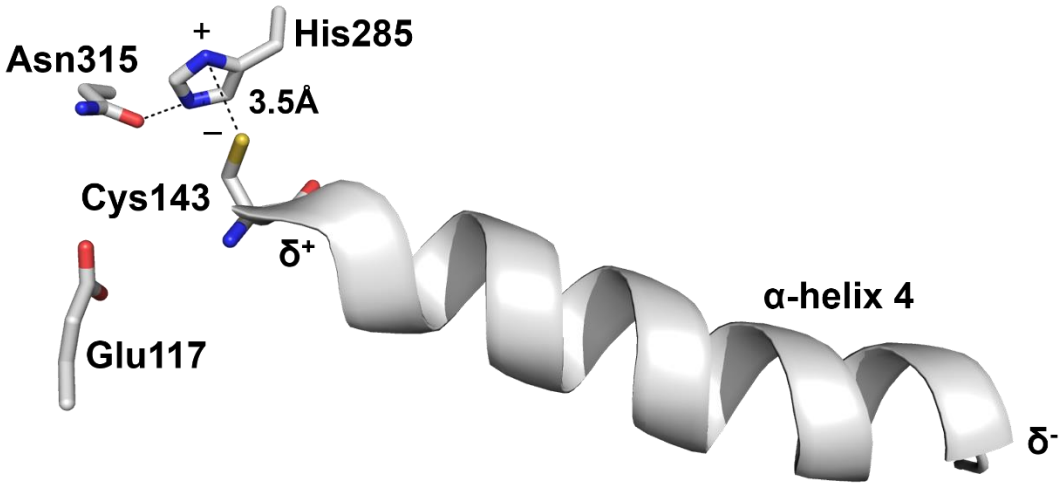


Figure 36: Position of Cys143 in line with the α-helix 4 dipole moment. WT OleA position of Cys143 in line with the α-helix 4 dipole moment (PDB 3ROW) (25). His285 shares a hydrogen bond with the side chain of Asn315 and is 3.5 Å from the thiol of Cys143 (black dashed lines).

Mutating His285 to Ala or Asn did not impact the ability to hydrolyze myristoyl-CoA compared to WT OleA (Table 9). In the absence of Cys143, catalytic futile hydrolysis of myristoyl-CoA occurs, but orders of magnitude slower than transesterification in WT OleA (54). As the hydrolysis rate is not impacted, it is likely that Cys143 in these variants is able to form the acyl-enzyme intermediate, further suggesting that His285 does not act directly in activating Cys143 for nucleophilic attack of the first substrate. The H285D OleA hydrolyzes myristoyl-CoA at a slower rate, likely due to a combination of Cys143 pK_a increase that decreases the amount of reactive thiolate and negative charge repulsion with the carboxylate of Asp285 leading to non-optimal positioning of activated Cys143.

None of the OleA His285 variants are able to form β-keto acid, indicating they are able to catalyze transesterification but are unable to undergo Claisen condensation (Table 10). This is similarly observed in other thiolases. Active site histidine mutants of HMG-CoA synthase catalyze the acetyl transfer half reaction (transfer acetyl group from acetyl-CoA to cysteine) at a similar or faster rate than WT HMG-CoA synthase, but the rate of condensation is greatly diminished (69). It was suggested that the active site histidine acts to selectively stabilize and position the second substrate, acetoacetyl-CoA for condensation (28,69). Likewise, H244A FabH is efficient in transacylation to Cys112 but is deficient in decarboxylation of the second substrate malonyl-CoA that initiates condensation (67). Taken together, it is reasonable to hypothesize that His285 is required to position substrate for efficient Claisen condensation.

Similar to HMG-CoA synthase and FabH, we propose His285 interacts with substrate to promote condensation. This interaction is likely through its role in forming half of oxyanion hole 1 in OleA. HMG-CoA synthase coordinates the acetyl oxygen of its acetylated-Cys111 intermediate in an environment equivalent to OleA oxyanion hole 2 (Figure 37A) (23). Similarly, FabH binds its acetylated-Cys112 in oxyanion hole 2 (Figure 37B) (78). In both enzymes, the binding of the second substrate is achieved through interaction with the active site histidine. His233 in HMG-CoA synthase has been shown to selectively interact with acetoacetyl-CoA to orient the substrate for nucleophilic attack by the Cys111-acetyl carbanion (23,69). This interaction occurs in the OleA-equivalent oxyanion

hole 1 (Figure 37A). FabH His244 and Asn274 have been shown to be necessary for the decarboxylation of malonyl-ACP in order to generate the nucleophile (21,67,78,81). This position is also equivalent to OleA oxyanion hole 1 (Figure 37B). We propose OleA uses these binding pockets in reverse to those of other thiolases, utilizing oxyanion hole 1 to position the first acyl-CoA and oxyanion hole 2 to position the second. As seen in Figure 37C, the thioester carbonyl oxygen of myristoyl-CoA in the C143A OleA crystal structure is bound in oxyanion hole 1. This state represents the OleA-substrate complex prior to transesterification. No crystal structure of acylated-OleA has been solved; however, this substrate complex likely demonstrates the binding mode of the first acyl-CoA. The first substrate bound in the novel alkyl channel A necessitates use of oxyanion hole 1 to position its carbonyl before and after transesterification, making it unavailable when the second substrate binds. Thus, the second substrate uses oxyanion hole 2 prior to condensation. This hypothesis fits with Glu117 acting as a general base to abstract a proton from C_2 of the second acyl-CoA (Chapter 3). Coordination of the alkyl chain of the second substrate in alkyl channel B would position the thioester carbonyl in oxyanion hole 2, placing C_2 ~3.5 Å from the carboxylate of Glu117 (based on substrate modelling in OleA, PDB code 4KU3) (Figure 37D). We hypothesize that in the absence of His285, oxyanion hole 1 is compromised, and the first substrate may instead utilize oxyanion hole 2, as proposed for HMG-CoA synthase and FabH. This would occlude the second substrate from binding in the proper orientation, preventing

Glu117 from activating it for carbon-carbon bond formation. This is consistent with the inability of these mutants to condense product.

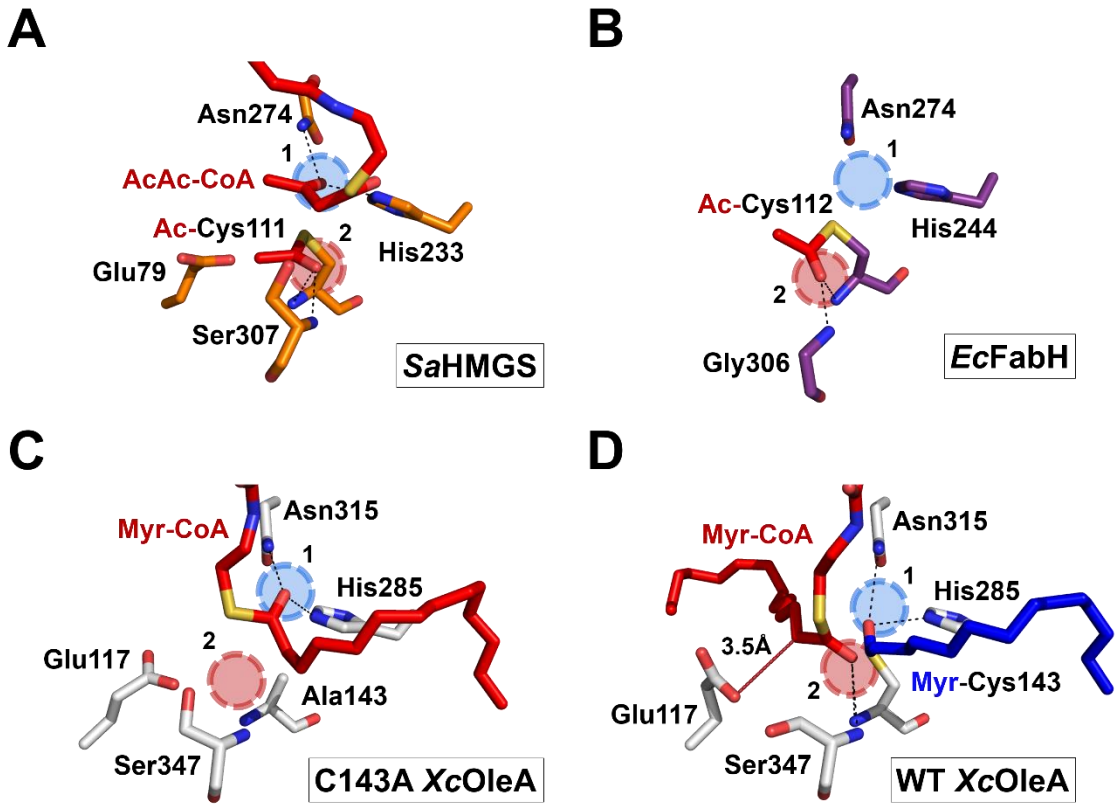


Figure 37: Spatial arrangement and usage of oxyanion holes in several thiolase enzymes. **A)** Acetylated HMG-CoA synthase (HMGS, orange) bound with acetoacetyl-CoA (AcAc-CoA, red) (PDB 1XPK) (23). The first substrate occupies oxyanion hole 2 (red circle) while the second substrate binds in oxyanion hole 1 (blue circle) **B)** Acetylated FabH coordinating substrate (red) in oxyanion hole 2 (red circle) (PDB 1HNH) (78). **C)** Enzyme-substrate complex between C143A OleA and myristoyl-CoA (Myr-CoA, red) (PDB 4KU2) (54). The carbonyl oxygen of Myr-CoA occupies oxyanion hole 1 (blue circle). **D)** WT OleA active site and substrate modelling displaying putative oxyanion hole usage before condensation. Myristic acid (blue) is acylated to Cys143 and occupies oxyanion hole 1 and alkyl channel A. Myristoyl-CoA (red) is bound in oxyanion hole 2 and alkyl channel B. Glu117 is modelled in the position found in OleA C143S (PDB 4KU3) and is located ~3.5 Å from the putative position of myristoyl-CoA C_2 atom (red line) (54). All models are superimposed and viewed in isolation from the same vantage point. Hydrogen bonds are shown as black dashed lines. *Sa, Staphylococcus aureus; Ec, Escherichia coli; Xc, Xanthomonas campestris.*

In summary, we report the crystal structures of H285A, H285N, and H285D OleA along with H285N OleA in complex with the inhibitor cerulenin. These structures reveal oxidative damage of Cys143 during X-ray data collection, as well as cleavage of the covalent bond between Cys143 and cerulenin. We propose His285 stabilizes the thiolate of Cys143 before transesterification and protects it from oxidative damage. In addition, we propose His285 is required to associate the transesterified acyl-Cys143 intermediate with oxyanion hole 1 such that the second substrate can bind productively in oxyanion hole 2 for deprotonation by Glu117 and initiate condensation. The binding of two long-chain alkyl chains during turnover in OleA dictates a mechanistic strategy that utilizes the oxyanion holes in reverse order compared to other thiolase enzymes.

CHAPTER 5

OleA catalytic mechanism conclusions

The work conducted in this dissertation has moved forward our understanding of the catalytic mechanism of OleA. Through site-directed mutagenesis of a series of active site amino acid residues, we have been able to assign and propose functional roles for each residue tested: Cys143, Glu117, and His285. Crystal structures for these mutant OleA enzymes have provided insights into structural rearrangements needed for binding substrate, necessary positioning of substrates for condensation, and potential interactions between amino acid side chains. It is our goal that the understanding of the events leading to condensation of two acyl-CoA substrates will aid future efforts to engineer OleA for tailored hydrocarbon production.

Conclusions from Cys143 investigation

Successful co-crystallizations of OleA C143A and C143S with myristoyl-CoA and lauroyl-CoA demonstrate the novel substrate binding architecture of OleA. Trapped substrate and intermediate complexes clearly define the binding mode of acyl-CoA in each of the three binding channels centered on Cys143, showing the hydrophobic alkyl chains bound in alkyl channels A and B and the CoA portion bound in the solvent-accessible pantetheinate channel. This spatial

configuration for coordinating two bulky substrates to diametrically opposed sides of the catalytic cysteine is a new feature for thiolase superfamily enzymes. The co-crystal structures also reveal features of *Xanthomonas campestris* OleA that allow it to bind and condense variable chain length acyl-CoAs. This is evident by the extended depth of alkyl channel A and the ability to accommodate different chain lengths in alkyl channel B by curling the alkyl chains back toward the active site. This is accomplished by polar and charged amino acid side chains located toward the back of alkyl channel B. These structures also reveal an important mechanistic detail in which alkyl channel A is the first channel to be bound by substrate.

Conclusions from Glu117 investigation

The invariant Glu117 residue in OleA is the first example of a thiolase containing an active site base originating from the opposite monomer of the dimeric enzyme. Through E117A, E117Q, and E117D variants and subsequent structural and biochemical analyses, this glutamate is demonstrated to be the general base necessary for initiation of Claisen condensation. The negative charge from the carboxylate of Glu117 is also suggested to help position the first substrate into alkyl channel A for transesterification. Based on structural arguments from Cys143 mutant co-crystals, an updated mechanism is proposed. Unlike the sequentially and structurally similar thiolase HMG-CoA synthase, OleA activates the second, non-covalently bound substrate for nucleophilic attack.

Glu117 acts by abstracting a proton from C_2 of the substrate bound in alkyl channel B in order to generate a carbanion that attacks to thioester of the acyl-Cys143 intermediate. This mechanistic strategy and three-channel arrangement suggest that OleA represents a novel hybrid subcategory of thiolase enzyme mechanisms.

Conclusions from His285 investigation

OleA His285 variants reveals enzymes that are competent in acyl-CoA hydrolysis but cannot form condensed product. Crystal structures of these variant enzymes show oxidative damage to Cys143 as a result of X-ray data collection. These are the only examples of OleA mutations that show susceptibility to oxidation, suggesting a protective interaction between His285 and Cys143. Even though cerulenin is able to covalently modify H285N OleA, the crystal structure of this complex shows that radiative damage has broken the carbon-sulfur bond. Because these His285 variants are likely able to complete transesterification but are unable to form product, an updated mechanism is proposed. The side chain of His285 forms half of oxyanion hole 1 alongside the side chain of Asn315. This oxyanion hole is thought to not only stabilize the negatively charged intermediate during transesterification but also position the thioester of the acyl-Cys143 intermediate prior to condensation. This anchoring of substrate allows the second substrate to bind in alkyl channel B and coordinate proton abstraction by Glu117 through substrate orientation to engage oxyanion hole 2.

Future directions

An outstanding question in the catalytic mechanism of OleA deals with the final hydrolysis step. Following condensation, a β-keto acyl-CoA species is formed, and CoA needs to be hydrolyzed by OleA before the β-keto acid can be reduced by OleD. This chemical species is implicated by our mechanistic model in which the second substrate attacks the acyl-Cys143 intermediate. The intermediate has been demonstrated by HPLC and mass spectrometry, and it was shown to be hydrolyzed by OleA (16). In Chapter 3 we propose that an activated water molecule is responsible for this hydrolytic step. However, we have recently prepared OleA mutants of a second conserved active site cysteine (Cys239) that exhibits a diminished rate of acyl-CoA hydrolysis. C239G and C239S OleA variants are able to form β-keto acid, suggesting the role of Cys239, if any, occurs after condensation. Based on its spatial proximity to the thioester of the second substrate in the C143S myristoyl-CoA co-crystal structure, Cys239 is a potential candidate for direct or indirect CoA hydrolysis. An experiment that could address the role of Cys239 would be to test the final hydrolysis directly. A chemical synthesis route to produce a β-keto acyl-CoA analogue (replaces CoA with N-acetyl cysteamine, SNAC) has been published and would be adapted to utilize substrate alkyl chain lengths suitable for *X. campestris* OleA (Figure 38) (16). This analogue would be reacted with WT and C239G/S OleA to determine the concentration of hydrolyzed SNAC by DTNB absorbance as well as β-keto acid formation by GC/MS. If Cys239 plays a role in hydrolyzing CoA from the

final product, this assay would show C239G/S mutants as deficient in hydrolysis and product formation.

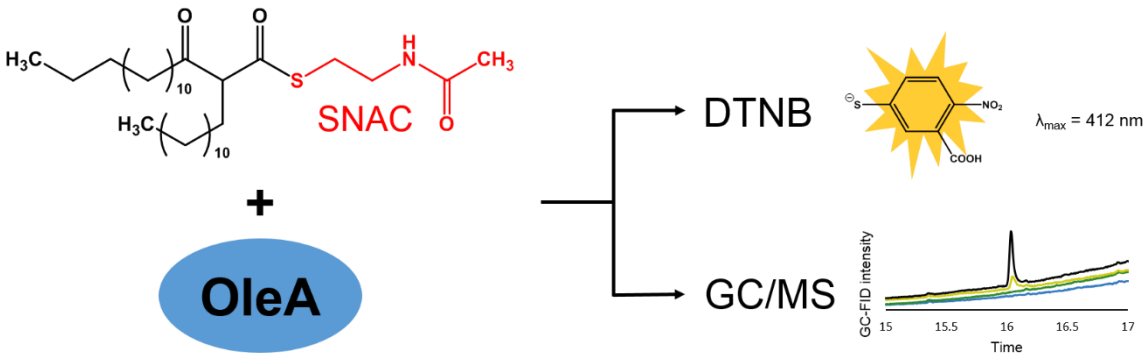


Figure 38: Experimental design to test hydrolysis of synthetic β-keto acyl-SNAC by WT and C239G/S OleA. Analysis by reaction with DTNB would measure the concentration of hydrolyzed SNAC. Analysis by GC/MS would detect hydrolyzed and thermally-decarboxylated β-keto acid.

The next step is to apply our acquired mechanistic knowledge to engineer OleA for specific condensation reactions. A good place to start would be to introduce mutations to control the types of acyl-CoA accepted by each alkyl binding channel. Natively, *X. campestris* OleA accepts C_8-C_{16} acyl-CoA, and controlling the length of alkyl chain that binds in each alkyl channel would lead to specific β-keto acids and ultimately specific olefins. In alkyl channel A, hydrophobic residues, such as Ile258 and Ala261, help guide the acyl chain deeper into the alkyl channel. If these residues were replaced with bulky or polar amino acids, the alkyl channel might only accept shorter chain length acyl-CoA. We observe that the first substrate binds in alkyl channel A for transesterification. This observation may prove helpful in determining the type of hydrocarbon condensed because the shorter chain substrate would be attacked by the

nucleophilic second substrate. If the opening to alkyl channel A could be completely occluded by amino acid substitution, reaction with acetyl-CoA as the first substrate might lead to the intriguing production of terminal olefins. Similarly in alkyl channel B, mutagenesis of residues positioned near the back of the channel could aid in selection of acyl-CoA chain length available for condensation.

CHAPTER 6

This chapter (with supplemental material reintegrated) is reprinted with permission from American Chemical Society *Biochemistry*.

Biochemistry, 2017, ***56***, pp 348–351 DOI: 10.1021/acs.biochem.6b01199

β-Lactone synthetase found in the olefin biosynthesis pathway

Christenson, J.K., Richman, J.E., Jensen, M.R., Neufeld, J.Y., Wilmot, C.M. and Wackett, L.P.

Author contribution

For the work presented in this chapter, Jensen, M.R. helped optimize purification conditions for OleC enzyme, purified recombinant OleC enzyme, and aided experimental design and implementation. Jensen, M.R. also contributed to the final editing of the paper.

Chapter 6 summary

The first β-lactone synthetase enzyme is reported, creating an unexpected link between the biosynthesis of olefinic hydrocarbons and highly functionalized natural products. The enzyme OleC, involved in the microbial biosynthesis of long-chain olefinic hydrocarbons, reacts with *syn*- and *anti*-β-hydroxy acid substrates to yield *cis*- and *trans*-β-lactones, respectively. Protein sequence

comparisons reveal that enzymes homologous to OleC are encoded in natural product gene clusters that generate β-lactone rings, suggesting a common mechanism of biosynthesis.

Manuscript

The β-lactone (2-oxetanone) substructure is well-known in organic synthesis and microbial natural products, some of which are presently being investigated for anti-obesity, anticancer, and antibiotic properties (82-86). Although multiple organic synthesis routes exist for β-lactones, no specific enzyme that catalyzes the formation of this functional group had been identified (87). While defining the chemistry of a well-known olefinic hydrocarbon biosynthesis pathway, we identified a β-lactone synthetase whose presence extends into natural product biosynthesis.

The olefin biosynthesis pathway is encoded by a four-gene cluster, *oleABCD*, and is found in more than 250 divergent bacteria (13). Ole enzymes produce long-chain hydrocarbon *cis*-alkenes from activated fatty acids through the reported chemistry shown in Figure 39 (11,88). OleA, the first enzyme of the pathway, has been studied in *Xanthomonas campestris* (*Xc*) and found to catalyze the “head-to-head” coupling of CoA-activated fatty acids (**1**) to unstable β-keto acids (**2**) (15). The second enzyme, OleD, couples the reduction of **2** with NADPH oxidation to yield stable β-hydroxy acids (**3**) as defined in *Stenotrophomonas maltophilia* (*Sm*) (16). Finally, using gas chromatography

(GC) detection methods, we have observed and others have reported that *Sm* OleC catalyzes an apparent decarboxylative dehydration reaction to generate the final *cis*-olefin product (19). Together, these findings indicate that there is no defined purpose for the ever-present fourth gene in the cluster, *oleB*.

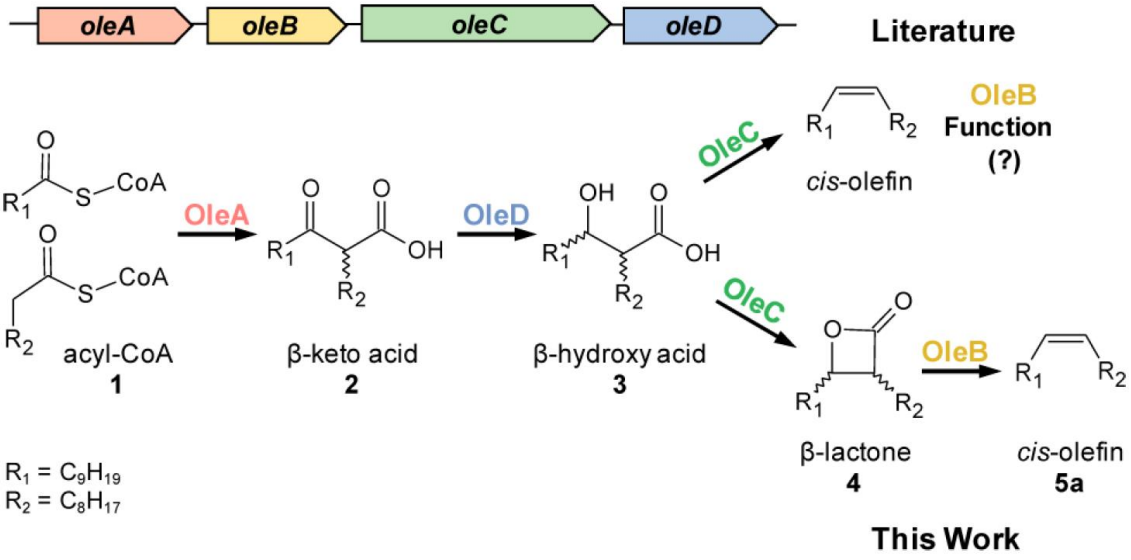


Figure 39: *ole* gene cluster and enzymatic pathway.

We now report that the observation of *cis*-alkenes as products of OleC is an artifact, a misleading observation caused by GC analytical methods used to identify the high-boiling point products of OleC. Rather, using ^{1}H nuclear magnetic resonance (NMR), we demonstrate that OleC proteins from four different bacteria produce thermally labile β-lactones from β-hydroxy acids in an ATP-dependent reaction; no alkenes are observed. Further analyses of gene clusters for β-lactone-containing natural products reveal *oleC* homologues that likely perform this previously unknown biological β-lactone ring closure reaction.

The first suggestions of β-lactone synthetase activity arose when monitoring reactions of *Xc* OleC with ATP, $MgCl_2$, and a synthetic, diastereomeric mixture of **3** by GC. Two peaks were observed by GC, coupled to both a mass spectrometer and flame ionization detector (FID), with mass spectra and retention times identical to those of synthetic *cis*- and *trans*-olefin standards (**5a** and **5b**, respectively). However, the GC/FID peak area of the enzymatically generated olefin varied significantly with GC inlet temperature and inlet liner purity, while synthetic standards were unaffected, suggesting that the observed olefin may be a thermal decomposition product of the actual OleC product.

To test this hypothesis, reactions of *Xc* OleC with **3** were scaled to generate sufficient quantities for ^{1}H NMR. No resonances consistent with the prepared olefin standards (**5a** and **5b**) were observed; rather, four distinct multiplets, each appearing as a doublet of doublets of doublets, arose. These resonances were consistent with the two hydrogens of *cis*- and *trans*-β-lactone rings and perfectly matched our authentic standards of *cis*- and *trans*-3-octyl-4-nonyloxetan-2-one (**4a** and **4b**, respectively). Furthermore, when **4a** and **4b** were analyzed by GC, retention times and mass spectra matched those of olefin standards **5a** and **5b**, with sensitivity to inlet conditions being observed. The thermal decarboxylation of *cis*- and *trans*-β-lactone to *cis*- and *trans*-olefin, respectively, is well known (89,90). We believe that thermal decomposition during GC/mass spectrometry (MS) analysis caused the product of OleC catalysis to be misidentified. Additionally, when supplemental NMR data from the

literature report of *Sm* OleC characterization were reviewed, resonances of the *cis*- and *trans*-β-lactones, consistent with those synthesized here, are visible (19).

The stereochemical origins of **4a** and **4b** were then investigated by reacting *Xc* OleC with *syn*- and *anti*-β-hydroxy acids. High-performance liquid chromatography was used to separate **3** into its *syn*- and *anti*-diastereomeric pairs (**3a** and **3b**, respectively) (Figure 40). Examining **3a** and **3b** by ^{1}H NMR and GC/MS, post-methylation, demonstrated each contained <10% of the opposite racemic diastereomer. When reacting with *Xc* OleC, **3a** produced **4a** while **3b** generated **4b** (Figure 41). GC/MS analysis supported this conclusion, as OleC reactions with **3a** and **3b** yielded the β-lactone breakdown products, **5a** and **5b**, respectively. OleC consumed >90% of substrates **3a** and **3b** as determined by GC/MS, supporting the conclusions of Kancharla et al. that all four β-hydroxy acid isomers are utilized by OleC (19). Taken together, *Xc* OleC represents the first reported β-lactone synthetase, converting β-hydroxy acid substrates to β-lactones in the presence of ATP and $MgCl_2$. MgATP is likely required to activate the hydroxyl or carboxyl group and promote β-lactone ring formation.

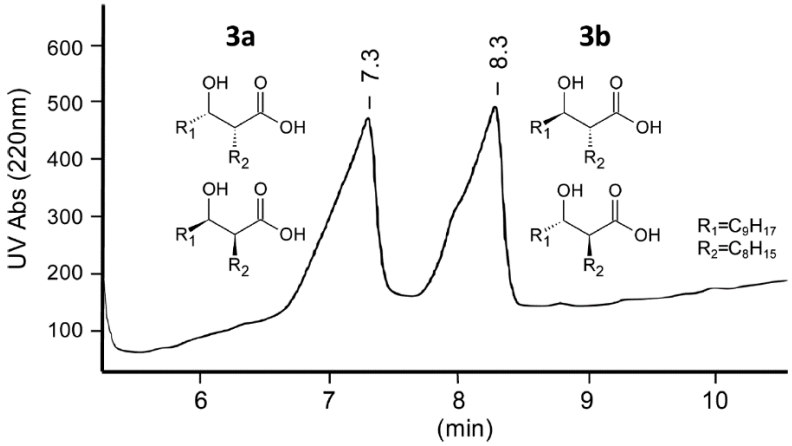


Figure 40: High-performance liquid chromatography separation of *syn*- and *anti*-β-hydroxy acids (**3a** and **3b**, respectively) for use as *Xc* OleC substrates. The rise in baseline is caused by impurities from the MTBE gradient that absorb at 220 nm.

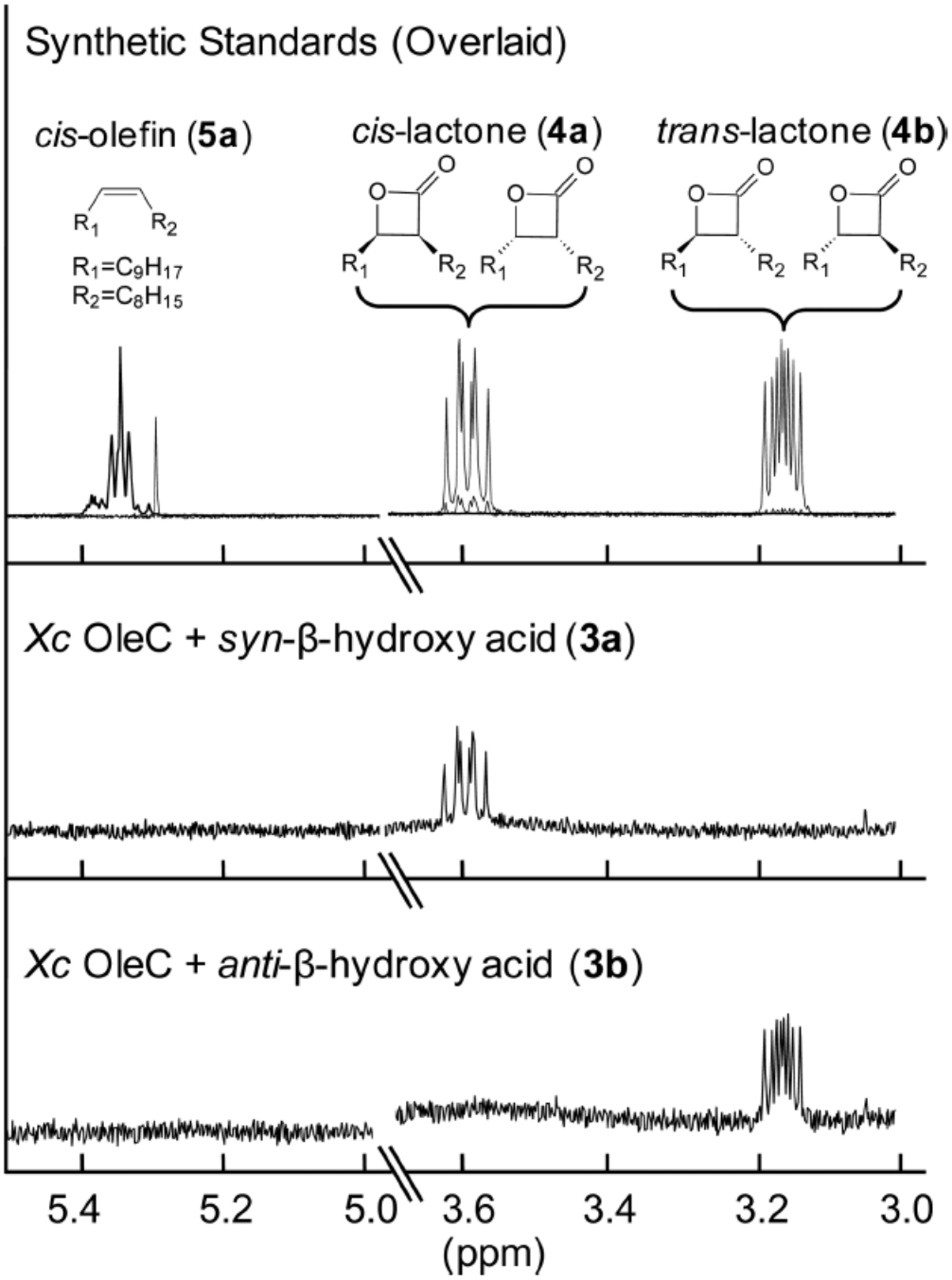


Figure 41: ^{1}H NMR (400 MHz, $CDCl_3$) analyses of *Xc* OleC products with β-hydroxy acid substrates.

To determine if β-lactone synthetase activity is a common enzymatic step in long-chain olefin biosynthesis, we obtained four *oleC* genes from *oleABCD*

gene clusters in divergent microorganisms (Table 12). Purified OleC enzymes from the four organisms were reacted overnight with ATP, $MgCl_2$, and **3** and then analyzed by ^1H NMR and GC/MS. The products of OleC proteins from the bacteria *S. maltophilia*, *Arenimonas malthae*, and *Lysobacter dokdonensis* were both **4a** and **4b** β-lactones, with no **5a** or **5b** olefins being observed, indicating that OleC enzymes from diverse sources are β-lactone synthetases. The Gram-positive bacterium *Micrococcus luteus* (*Ml*) was specifically chosen because its sequence diverges greatly from that of *Xc* OleC, and it contains a natural fusion of the *oleB* and *oleC* genes. This natural *oleBC* fusion is found in actinobacteria, which comprise ~30% of the microorganisms that contain identifiable oleABCD genes. Reaction of the purified *Ml* OleBC fusion with $MgCl_2$, ATP, and **3** produced β-lactones **4a** and **4b** as well as small amounts of *cis*-olefin, **5a**. No trace of *trans*-olefin, **5b**, was detected. Further characterization is ongoing, but we believe that OleB performs a *syn* elimination of carbon dioxide from the *cis*-β-lactone to form the final *cis*-olefin product. This is consistent with previous studies of microorganisms expressing ole genes that contain olefins with a *cis* relative configuration exclusively (11,13,88). These data also demonstrate that an enzyme domain with an amino acid sequence only 35% identical to that of the *Xc* OleC generates β-lactones, indicating that this activity is likely common among all olefinic hydrocarbon biosynthesis OleC homologues.

Table 12: Other OleC enzymes make β-lactones.

Organism	Accession no.	% ID[a]
X. campestris	WP_011035474.1	100
S. maltophilia	AFC01244.1	77
A. malthae	WP_043804215.1	73
L. dokdonensis	WP_036166093.1	70
M. luteus[b]	WP_010078536.1	35

[a]Percentidentity based on amino acid sequence. [b]OleC and OleB are a natural fusion in *Ml.*

Establishing the widespread nature of lactone synthetase activity within olefinic hydrocarbon biosynthesis led us to search sequence databases for OleC homologues in other biosynthetic pathways. OleC is a member of the ubiquitous AMP-dependent ligase/synthetase enzyme superfamily; as such, homologues are found in all organisms (91). As of November 2016, a BLAST search of NCBI's nonredundant protein sequence database identified more than 900 sequences with >35% sequence identity and more than 16000 with >25% sequence identity to *Xc* OleC.

Of the sequences examined, two *Xc* OleC homologues were clearly encoded in gene clusters known to produce β-lactone natural products (Figure 42). The first, LstC, is an uncharacterized enzyme found in the lipstatin biosynthesis pathway from *Streptomyces toxytricini*. Lipstatin is the precursor to Orlistat, the only over-the-counter, Food and Drug Administration-approved anti-obesity drug. LstC is a member of the AMP-dependent ligase/synthetase superfamily, and its protein sequence is 38% identical to that of *Xc* OleC, more similar than the sequence of the β-lactone synthetase domain of *Ml* OleBC (35%). Surprisingly, further investigation revealed homologues of OleA and OleD

encoded by the lipstatin gene cluster, suggesting that the two gene clusters have a common ancestry. The syntheses of both lipstatin and olefinic hydrocarbons are initiated by the condensation of two fatty acyl- CoAs to form a β-keto acid. In the case of lipstatin, the two fatty acids are 3-hydroxy-linoleic and octanoic acid (82). The hydroxyl group of linoleic acid is later functionalized by LstE and LstF with a modified valine (82). LstD and OleD likely perform the same NADPH-dependent reduction of the β-keto group to a hydroxyl group. Formation of the *trans*-β-lactone is likely accomplished by the OleC homologue LstC, to generate the final product in lipstatin biosynthesis. Olefin biosynthesis is completed by the putative OleB-dependent elimination of CO_2 to generate the final olefin product. The lipstatin gene cluster lacks any gene product that is homologous to OleB, consistent with the accumulation of the β-lactone natural product and further supporting our hypothesis that OleB performs the final step in the biosynthetic pathway to olefins.

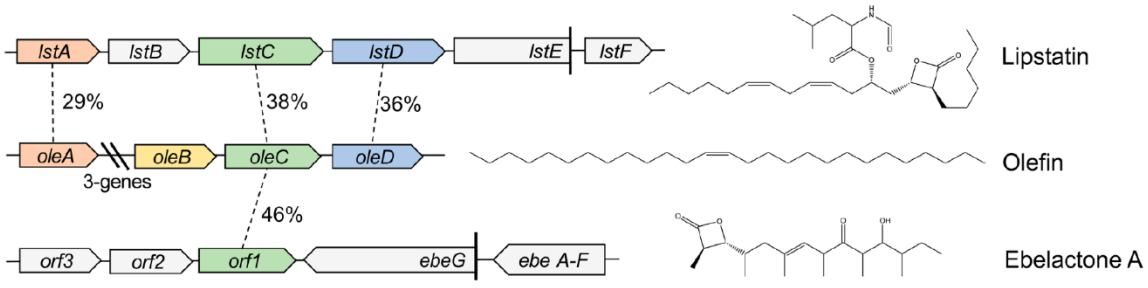


Figure 42: Homologous OleC enzymes encoded in β-lactone biosynthesis gene clusters. Percent identity is based on amino acid sequences. The *E* values for OleC to LstC and Orf1 are 2×10^{-72} and 1×10^{-143}, respectively. Lipstatin is the precursor to the anti-obesity drug Orlistat. Ebelactone A is a commercially available general esterase inhibitor.

The gene cluster responsible for the biosynthesis of ebelactone A, a commercially available esterase inhibitor, in *Streptomyces aburaviensis* shows a gene, *orf1*, with an amino acid sequence 46% identical to that of *Xc* OleC and is directly adjacent to *ebeA−G* (Figure 42). Unlike lipstatin, ebelactone A is formed partly by a polyketide synthase multidomain protein rather than fatty acid condensation; as such, OleA and OleD homologues are not encoded in the surrounding gene cluster. Literature reports suggest that the β-lactone ring of ebelactone A is formed spontaneously from the final, enzyme-linked, β-hydroxy-thioester intermediate (86). While a spontaneous β-hydroxy-thioester cyclization is mechanistically plausible, β-lactone formation from β-hydroxy-thioesters in ubiquitous pathways such as fatty acid oxidation or synthesis has not been reported to the best of our knowledge (92). Additionally, β-hydroxy-thioester intermediates are extremely common in polyketide synthesis pathways, while β-lactone formation is comparatively rare. An Orf1-independent cyclization would require a unique property of ebelactone A precursors or a novel polyketide domain architecture to promote β-lactone ring cyclization. However, in favor of an Orf1-independent mechanism is the fact that no thioesterase domain exists in the final polyketide synthesis domain, suggesting that no free β-hydroxy acid is released for the putative ATP-dependent Orf1 to act on. Other polyketide-type β-lactone gene clusters, such as those for salinosporamide A, cinnabaramide, and oxazolomycin, do not encode an OleC homologue with high sequence identity (>35%) in the vicinity of the cluster. Polyketide-derived β-lactones are thought to

form by the cyclization of the final thioester, enzyme-linked intermediate, but this has never been characterized (83,93-95). It is reasonable to hypothesize that specialized polyketide synthase domains represent a second mechanism of β-lactone formation. Regardless, the discovery of a stand-alone β-lactone synthetase here creates new opportunities for the natural product field. Preliminary screening of *Streptomyces* and *Nocardia* genomes suggests that β-lactone natural products may be more widespread than currently realized.

Materials and methods

Synthesis and characterization of β-hydroxy acids and β-lactones

The synthesis, isolation, and characterization of compounds **3a**, **3b**, **4a**, **4b**, **5a**, and **5b** were carried out by Dr. Jack E. Richman. **3** was synthesized and separated into its *syn-* and *anti-* diastereomeric pairs (**3a** and **3b**) using high performance liquid chromatography (HPLC). All products were verified by ^{1}H NMR and GC/MS (spectra not shown).

^{1}H NMR analyses of the Ole-pathway metabolites

All NMR spectra were acquired on a Varian Inova 400 MHz NMR spectrometer using a 5 mm Auto-X Dual Broadband probe at 20°C. Typically 512 pulses were used with a one second pulse delay. Solvent was always $CDCl_3$ with TMS as a reference. Enzyme reactions contained 1-bromo-napthalene as an integration reference. Common known contaminant resonances in our spectra are the $CDCl_3$ (singlet, 7.26 ppm), dichloromethane (singlet, 5.30 ppm), and the

plasticizer bis(2-ethylhexyl) phthalate (dioctyl phthalate) (multiplet 4.22 ppm). The origin of the quartet at 3.98 ppm in several figures is unknown.

GC/MS analysis of Ole pathway metabolites

Separation and identification of metabolites was accomplished by gas chromatography/mass spectrometry (GC/MS Agilent 7890a & 5975c) equipped with an Agilent J&W bd-ms1 column (30 m, 0.25 mm diameter, 0.25 μm film). Olefin (from either β-lactone breakdown or standards) was detected without derivatization, but the β-hydroxy acid required methylation of the carboxylic acid group by diazomethane. Preparation of ethereal alcoholic solutions of diazomethane from N-methyl-N-nitroso-*p*-toluenesulfonamide was carried out using a commercial kit from Aldrich (Technical information Bulletin No. AI-113). All samples were extracted with ethyl acetate (EtOAc) and mixed with 100 μL of diazomethane solution when needed. Five microliters of each sample were injected into the injection port (230° C). The 20 min program was as follows: hold 100° C for 3 min; ramp linearly to 280° C for 16 min; hold 280° C for 2 min.

Protein expression and purification

All genes were *Escherichia coli (E. coli)* codon optimized and obtained from either IDT or DNA 2.0. Constructs were created using standard cloning techniques with restriction digest, ligation, and transformation reagents all purchased from New England Biolabs (NEB) and used according to manufacturer's protocols. Constructs with pET28b^{+} contained an N-terminal 6X

His tag, while pET30b$^+$ contained a C-terminal tag. All vector constructs were sequence verified and are listed in Table 13.

Table 13: Vector constructs used in this study.

Protein	Organism	Accession no.	Vector	Buffer additive
OleA	*Xanthomonas campestris*	WP_011035468.1	pET28b$^+$	5 mM BME
OleC	*Xanthomonas campestris*	WP_011035474.1	pET30b$^+$	–
OleD	*Xanthomonas campestris*	WP_011035474.1	pET28b$^+$	0.025% Tween 20
OleC	*Stenotrophomonas maltophilia*	AFC01244.1	pET30b$^+$	–
OleC	*Arenimonas malthae*	WP_043804215.1	pET30b$^+$	–
OleC	*Lysobacter dokdonensis*	WP_036166093.1	pET30b$^+$	–
OleB-C	*Micrococcus luteus*	WP_010078536.1	pET30b$^+$	–

OleB-C is a natural fusion of OleB and OleC in *Ml*.

All vectors were transformed into BL21 (DE3) cells (NEB). Starter cultures (5 mL) were grown in LB media overnight at 37° C with kanamycin selection. Cells were transferred to 2 L flasks containing 1 L of LB media with kanamycin and grown to an optical density of 0.5 at 37° C. Growth was transitioned to 16° C, and expression was induced by isopropyl β-D-1-thiogalactopyranoside (IPTG, 100 μM). Cells were pelleted and resuspended in 10 mL of buffer containing 200 mM NaCl, 20 mM $NaPO_4$, and 10% glycerol at pH 7.4. Protein-specific buffer additives are indicated in Table 13. Cells were sonicated, spun at 33,000 x *g* for 45 min, and the supernatant was filtered through 0.45 and 0.22 μm filters (Millipore), successively. The supernatant was then loaded onto a GE HisTrip HP 1 mL column using an ÄKTA fast protein liquid chromatography instrument (GE). The column was first washed with 125 mM imidazole buffer before protein elution with 400 mM imidazole buffer. Protein concentration was determined using the

Bradford protein assay before samples were aliquoted and frozen at -80° C for future use.

Formation of β-lactones by OleC from β-hydroxy acids

Reactions to generate β-lactones contained 1.0 mg OleC, 20 mg of ATP, and 30 mg of $MgCl_2•6H_2O$ in 100 mL of buffer without glycerol (200mM NaCl and 20 mM $NaPO_4$ at pH 7.4). The appropriate β-hydroxy acids were dissolved in DMSO to 1.0 mg/mL, and 1-bromo-napthalene at 0.5 mg/mL was included as an internal reference. Replacing DMSO with ethanol for substrate delivery had no effect on the results. Reactions were initiated with 1.5 mL of the DMSO-substrate solution and allowed to run for 24 hours before three successive extractions were performed with 10 mL, 5 mL, and 5 mL of dichloromethane. Samples were allowed to evaporate at room temperature before dissolving in $CDCl_3$ for NMR. Extraction with ethylacetate instead of dichloromethane produced the same results.

Computational analyses of OleC homologues

A BLAST search with the non-redundant protein sequence database from the National Center for Biotechnology Information (NCBI) was used to identify homologs of OleC (WP_011035474.1) from *Xanthomonas campestris pv. campestris* ATCC 33913. BLAST parameters were as follows: matrix = BLOSUM62; gap cost = Existence:11 Extentension:1; word size = 6; max matches in a query range = 0; expect threshold = 10. The most distantly related sequences that maintained 25% identity had E values of $2e^{-14}$ or better.

CHAPTER 7

Active multienzyme assemblies for long-chain olefinic hydrocarbon biosynthesis

Christenson, J.K., Jensen, M.R., Goblirsch, B.R., Mohamed, F.A., Zhang, W., Wilmot, C.M., Wackett, L.P.

Author contribution

For the work presented in this chapter, Jensen, M.R. helped optimize purification conditions for OleBCD enzyme assemblies, purified recombinant OleBCD enzyme assemblies, aided experimental design and implementation, and performed all transmission electron microscopy experiments with the assistance of Zhang, W. Jensen, M.R. also helped draft portions of the manuscript and contributed to the final editing of the paper.

Chapter 7 summary

Bacteria from different phyla produce long-chain olefinic hydrocarbons derived from an OleA-catalyzed Claisen condensation of two fatty acyl coenzyme

A (acyl-CoA) substrates, followed by reduction and oxygen elimination reactions catalyzed by the proteins OleB, OleC, and OleD. In this report, OleA, OleB, OleC, and OleD were individually purified as soluble proteins, and all were found to be essential for reconstituting hydrocarbon biosynthesis. Recombinant coexpression of tagged OleABCD proteins from *Xanthomonas campestris* in *Escherichia coli* and purification over His_6 and FLAG columns resulted in OleA separating, while OleBCD purified together, irrespective of which of the four Ole proteins were tagged. Hydrocarbon biosynthetic activity of copurified OleBCD assemblies could be reconstituted by adding separately purified OleA. Immunoblots of nondenaturing gels using anti-OleC reacted with *X. campestris* crude protein lysate indicated the presence of a large protein assembly containing OleC in the native host. Negative-stain electron microscopy of recombinant OleBCD revealed distinct large structures with diameters primarily between 24 and 40 nm. Assembling OleB, OleC, and OleD into a complex may be important to maintain stereochemical integrity of intermediates, facilitate the movement of hydrophobic metabolites between enzyme active sites, and protect the cell against the highly reactive β-lactone intermediate produced by the OleC-catalyzed reaction.

Importance

Bacteria biosynthesize hydrophobic molecules to maintain a membrane, store carbon, and for antibiotics that help them survive in their niche. The hydrophobic compounds are often synthesized by a multidomain protein or by

large multienzyme assemblies. The present study reports on the discovery that long-chain olefinic hydrocarbons made by bacteria from different phyla are produced by multienzyme assemblies in *X. campestris*. The OleBCD multienzyme assemblies are thought to compartmentalize and sequester olefin biosynthesis from the rest of the cell. This system provides additional insights into how bacteria control specific biosynthetic pathways.

Introduction

Long-chain olefinic hydrocarbons are abundant in nature. First described in 1929 in cabbage leaf extracts, plant C_{29} ketones and hydrocarbons were proposed to arise from a head-to-head condensation of fatty acids, with a loss of one of the “head” carboxylic acid groups as carbon dioxide (66). Bacterial long-chain olefinic hydrocarbons were characterized structurally in 1969 during studies on the lipid components of *Micrococcus luteus* (88). The genes and enzymes responsible for the head-to-head mechanism of hydrocarbon biosynthesis in bacteria were elucidated some 40 years later by the company LS9, Inc. (now Renewable Energy Group), and patented (10). A subsequent study in 2010 reported the presence of the four hydrocarbon biosynthetic genes, *oleABCD*, in 69 prokaryotes from many deeply rooted branches of the prokaryotic tree of life (Figure 43) (13). The cellular function(s) of the *ole* gene cluster has been studied only in *Shewanella oneidensis* MR-1. The presence of the *oleABCD* genes was found to promote more rapid cell growth during a shift to

colder temperatures, consistent with those genes being commonly found in cold-temperature bacteria from polar and marine environments (7,13,96). However, bacteria containing *ole* gene clusters inhabit a diversity of ecological niches, suggesting there may be a range of functions for long-chain olefinic hydrocarbons.

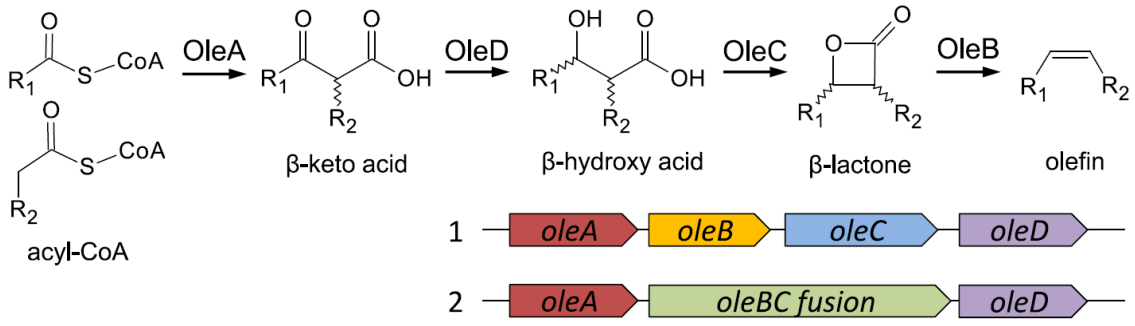


Figure 43: Olefin biosynthesis enzymatic pathway and gene clusters. Gene clusters 1 and 2 are the two most common gene arrangements found in *oleABCD*-containing organisms. The two fatty acyl-CoAs are typically linear alkyl chains between 10 and 16 carbons in length.

Only recently have the activities of each Ole enzyme been assigned (Figure 43), as initial work was hampered by the instability and hydrophobicity of the pathway intermediates. OleA is a soluble protein that condenses C_{10}-C_{16} fatty acyl groups via a nondecarboxylative Claisen condensation reaction to form a β-keto acid (15). If the β-keto acid is not quickly reduced by OleD, it spontaneously decarboxylates in aqueous solution to a ketone, as observed in earlier studies (12,15). Physiologically, OleD catalyzes an NADPH-dependent reduction of the β-keto acid intermediate to produce a β-hydroxy acid that is significantly more stable than the OleD substrate (16). Until recently, OleC was thought to catalyze the final reaction in olefin biosynthesis. However, OleC is now known to react

with the product of the OleD reaction to generate a thermally labile β-lactone (50). The β-lactone undergoes spontaneous and nonbiological decarboxylation to an olefin when monitored by gas chromatography (GC), which led to the initial incorrect assignment of OleC function (19,50). Recent data suggest that OleB acts as an unprecedented β-lactone decarboxylase to yield the final olefin product (50). Homologous gene clusters lacking an *oleB* gene have been identified in *Streptomyces*, and these produce β-lactone natural products rather than olefins (50). Microbial secondary metabolites containing β-lactones often serve as antibiotics and are known as general esterase and protease inhibitors (82,86).

It was considered here that Ole enzymes might form a multienzyme complex to effectively process the hydrophobic, unstable, and potentially toxic intermediates of olefin biosynthesis. The presence of an *oleBC* gene fusion in actinobacterial *ole* gene clusters further suggested a physical interaction among the enzymes of olefin biosynthesis. In the present study, we report the purification of active protein assemblies consisting of OleB, OleC, and OleD proteins. Assemblies of OleBCD were obtained by recombinantly coexpressing all four *ole* genes from *Xanthomonas campestris* pv. *campestris* ATCC 33913 (*X. campestris*) in *Escherichia coli* with various affinity tag combinations. Nondenaturing gels immunoblotted with anti-OleC showed OleC from an *X. campestris* cell lysate migrated as a high-molecular-weight band, similarly to purified OleBCD assemblies. The results from size exclusion chromatography

and electron microscopy were consistent with ordered assemblies of OleBCD, with an average molecular mass of ~1.9 MDa.

Results

Individual purification and physical properties of the four OleABCD enzymes

The purification of individually expressed OleA, OleC, and OleD proteins derived from different bacteria had been reported previously (7,12,15,16,19,50), while the purification of an OleB protein is reported here for the first time, to our knowledge. In this study, all four *X. campestris* Ole proteins were purified in a recombinant form from separate *E. coli* expression host cell lines. In contrast to conclusions reached in a previous study (19), we found that OleB is required for the reconstitution of hydrocarbon biosynthesis, consistent with the ubiquitous occurrence of *oleB* in the olefin gene cluster. The OleA and OleD reactions were previously demonstrated to produce a β-keto acid and β-hydroxy acid, respectively (Figure 43) (15,16). OleC, purified from *Stenotrophomonas maltophilia*, was reported to produce an olefin (19), but we recently demonstrated that OleC from that organism produces a β-lactone and that OleB is required to transform that to the final olefinic hydrocarbon (50).

Relevant physical properties of the Ole proteins are reported in Table 14. The oligomeric state of each protein was determined by gel filtration chromatography. However, detergents were required in the purifications of OleB and OleD, which could affect the observed oligomeric states. Moreover, all

attempts to concentrate either protein resulted in immediate precipitation, even in the presence of detergents. Glycerol increased the purification yields of OleB, OleC, and OleD to the levels shown in Table 14 but did not boost the yield of OleA. Since the proteins all appeared to be in the soluble fraction of the cell, the insoluble nature of OleB and OleD in the absence of detergents suggested to us that Ole proteins might associate with each other *in vivo*.

Table 14: Physical properties of the *X. campestris* Ole proteins.

	Value or characteristic for:			
Property	**OleA**	**OleB**	**OleC**	**OleD**
Subunit molecular mass (kDa)	36.6	34.1	58.5	36.1
Oligomeric state	Dimer	Tetramer	Monomer	Aggregate[a]
Required detergent		Triton X-100		Tween 20
Yield (mg/liter of culture)	15	1.9	5.8	1.1
Precipitated on concn?[b]	No	Yes	No	Yes

[a]*X. campestris* OleD runs at >600 kDa on SEC but remains active and soluble. *Stenotrophomonas maltophilia* OleD is a dimer (16).
[b]Precipitation of the sample after spin concentration in 30,000-MWCO Amicon spin filters.

OleBCD form active multienzyme assemblies without OleA.

To examine our hypothesis that Ole enzymes form a multienzyme complex, we simultaneously expressed *oleABCD* in *E. coli* with different affinity tag arrangements (1 to 3) shown in Figure 44A. Tag arrangement 1 contained an N-terminal His_6 tag on OleB and a C-terminal FLAG tag on OleC (His_6-OleBADC-FLAG). Even when no detergent was used, three clear SDS-PAGE bands corresponding to the size of OleC, OleA, and/or OleD, as well as OleB eluted from the Ni^{2+} column when the imidazole concentration reached ~140 mM. OleA and OleD are nearly indistinguishable by SDS-PAGE, as they differ by only 508

Da. Subsequent purification over an anti-FLAG column further purified these three bands. However, the addition of myristoyl coenzyme A (myristoyl-CoA) substrate and cofactors generated no olefin product unless separately purified OleA was added back to the reaction mix. The reaction products of OleA with copurified OleBCD were found to be identical to our synthetic *cis*-olefin standard (Figure 45). These data suggested that the association between OleBCD survives the purification, whereas OleA is lost during chromatography. For comparison, the individually expressed OleA, OleB, OleC, and OleD were reconstituted and also shown to produce olefin from CoA-charged fatty acids. In both samples, the absolute configuration of the major product was *cis*-olefin. No appreciable difference in the rate of olefin production was observed, possibly because the rate-limiting step for *in vitro* assays is release of the hydrophobic olefin from OleB.

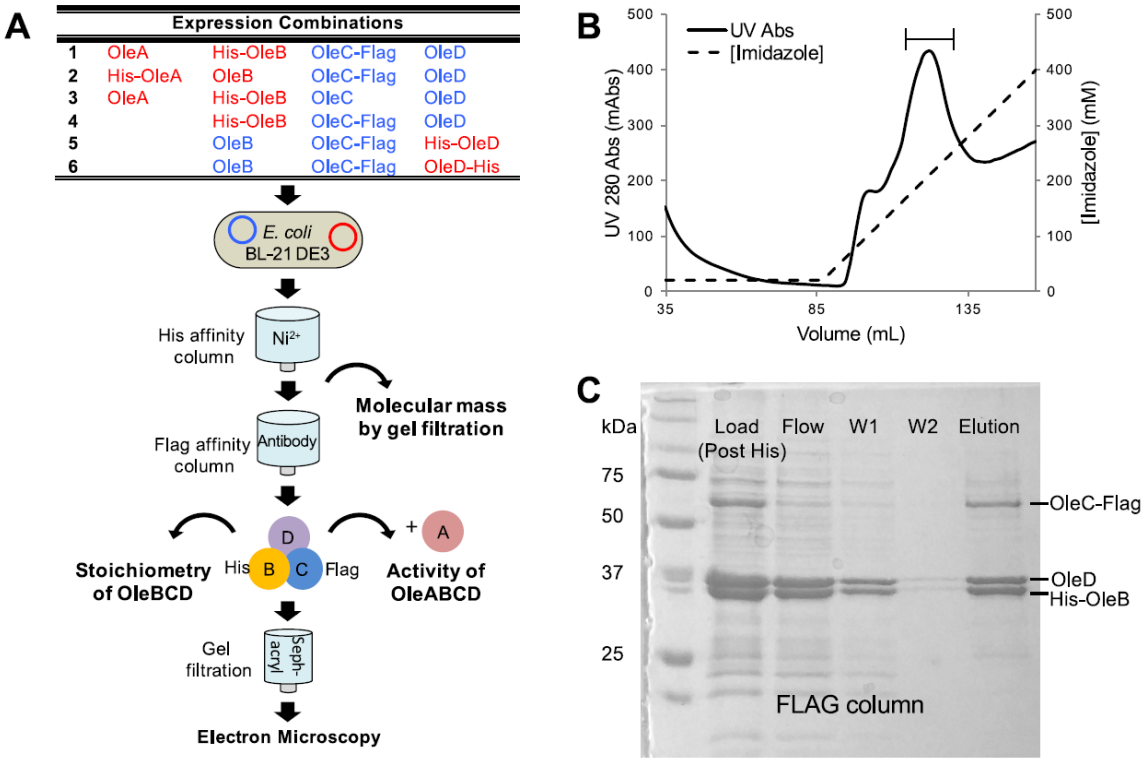


Figure 44: Plasmid expression, purification, and SDS-PAGE analysis of Ole protein coexpression. **A)** The table represents all tag combinations constructed. The placement of the His_6 and FLAG tags on the left and right of the protein corresponds to N- and C-terminal tags, respectively. Colors in the table represent the grouping of genes on a single vector for expression. **B)** Purification of OleBCD (tag arrangement 4) over His affinity column with increasing amounts of imidazole. Peak fractions indicated by the bar were collected, concentrated, and loaded onto the anti-FLAG column (lane labeled “Load” on gel in panel C). **C)** SDS-PAGE showing the post-His affinity column concentrate and the anti-FLAG purification of OleBCD tag arrangement 4. Bound samples were washed twice (W1 and W2) with buffer before elution with FLAG peptide. Note that OleC, the weakest-intensity band, contained the FLAG tag. Abs, absorbance.

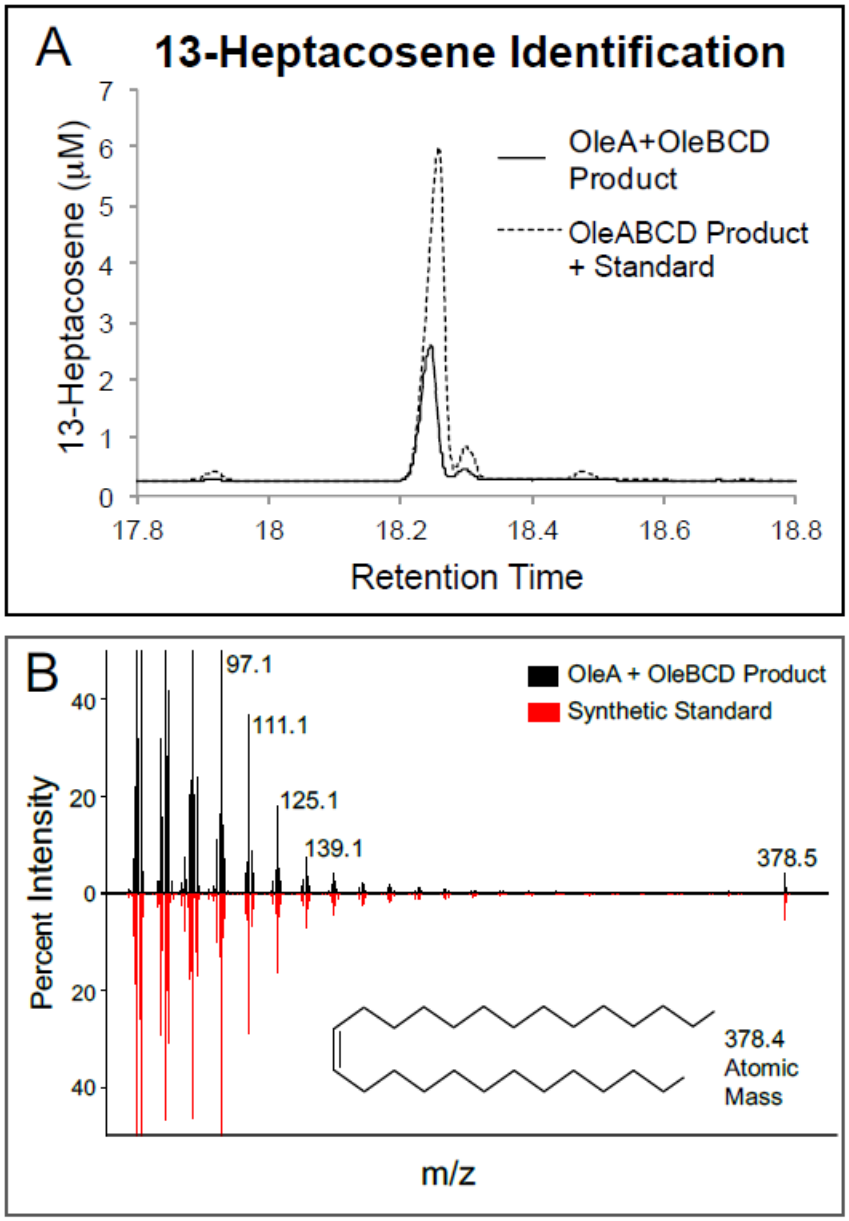


Figure 45: Confirmation of the olefin product by GC/MS. **A)** Flame Ionization Detector (FID) signal of the reaction product of Ole enzymes with myristoyl-CoA and the same reaction co-injected with synthetic cis-13-heptacosene. **B)** Mass spectrometric positive ion signal of the Ole enzyme product and the cis-13-heptacosene synthetic standard.

To examine if the tag placement altered OleA's binding to the putative complex of OleBCD, expression of tag arrangements 2 and 3 was undertaken.

Tag arrangement 2 (His_6-OleABDC-FLAG) resulted in the coelution of three bands at ~100 mM imidazole from the Ni^{2+} column that were identified by mass spectrometry (MS) as OleB, OleC-FLAG, and OleD, while His_6-OleA (MS identified) eluted separately at 200 mM imidazole. The observation that some OleBCD was retained on the Ni^{2+} column until imidazole was added suggested an interaction with His_6-OleA, which was disrupted prior to His_6-OleA being competitively eluted off the column. Tag arrangement 3 (His_6-BACD) resulted in the coelution of OleBCD with no OleA activity. Taken together, these data suggest that OleBCD form active stable multienzyme assemblies with which OleA weakly associates.

Optimizing the copurification of OleBCD.

To obtain pure assemblies of OleBCD routinely, OleA was no longer coexpressed with OleBCD; only *oleBCD* genes were expressed in the *E. coli* host. Of the three additional tag arrangements tested (4 to 6, Figure 44A), only the purification of tag arrangement 4, His_6-BDC-FLAG, yielded active assemblies after Ni^{2+} and anti-FLAG columns. In contrast to observations made with OleB and OleD proteins purified independently, the copurified OleBCD did not require detergents and could be concentrated without precipitation. A His_6 tag on OleD (either N- or C-terminal, tag arrangements 5 and 6) appeared to prevent formation of the enzyme assemblies. The purification of tag arrangement 4 is shown in Figure 44B and C. Glycerol was found to increase purification yields but was not required to obtain active assemblies. The individual components of this

OleBCD complex were confirmed by MS, and this sample was used in further characterization studies.

OleBCD complex identification in native X. campestris.

The level of expression of olefinic hydrocarbons in wild-type bacteria is modest (7,13), and so we used sensitive immunoblotting methods and nondenaturing gels to identify the anticipated multienzyme assembly within the native *X. campestris* host. A polyclonal antibody against *X. campestris* OleC was raised against a specific peptide predicted to be on the protein surface, as described in Materials and Methods. The polyclonal antibody was then validated by showing binding activity in immunoblots against purified OleC, copurified OleBCD, and lysates of *E. coli* expressing OleBCD. Figure 46 shows a nondenaturing gel that had been immunoblotted in which a lysate fraction derived from *X. campestris* cells is compared with copurified OleBCD and OleC alone. The band from the supernatant of *X. campestris* lysate migrates similarly to the OleBCD assemblies and more slowly than OleC alone. These data are consistent with the idea that the assemblies of OleBCD that form in *E. coli* likely assemble in a manner similar to that in the native host, *X. campestris*. It was also found in the soluble fraction, similar to recombinant expression in *E. coli*.

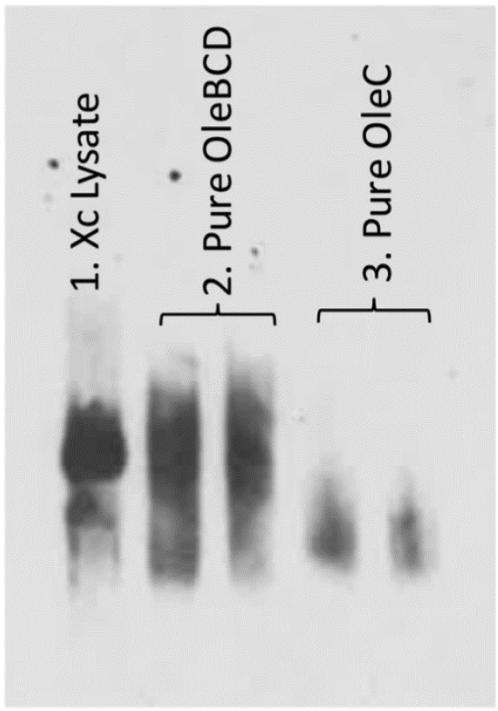


Figure 46: Anti-OleC immunoblot on a nondenaturing protein gel. Native OleC from the supernatant of lysed *X. campestris* cells migrates closely with recombinantly expressed and purified *X. campestris* OleBCD rather than individually purified OleC. Xc, *X. campestris*.

Size estimation of OleBCD assemblies by gel filtration chromatography.

Gel filtration chromatography was used to obtain insight into the molecular size of copurified OleBCD. Ni^{2+} column-purified OleBCD (tag arrangement 4) eluted in the void of a GE 16/600 Superdex 200 gel filtration column, indicating a size larger than 600 kDa. The void fraction was collected and found to be active upon addition of OleA, cofactors, and myristoyl-CoA. We subsequently attempted to purify OleBCD assemblies with detergents to see if a minimal functional complex could be isolated. However, OleBCD still ran in the void with the addition of Tween 20 or Triton X-100, while 3-[(3-cholamidopropyl)-

dimethylammonio]-1-propanesulfonate (CHAPS), a zwitterionic detergent, led to complete dissociation of OleBCD at all concentrations.

Subsequently, OleBCD was analyzed on a Sephacryl gel filtration column that is able to separate globular proteins up to 8 MDa. OleBCD (tag arrangement 4) eluted significantly after the void (45 ml) at a volume of 65.0 ml, consistent with a high molecular weight (Figure 47). The observed peak lies outside the range of commercially available protein standards (up to 600 kDa), but a molecular mass centered at ~1.9 MDa was determined by extrapolation. The activity of each fraction was assayed in the presence of excess OleA and activity tracked precisely with the A_{280}, consistent with stable active assemblies of OleBCD (Figure 47). SDS-PAGE analysis of fractions showed peak OleBCD protein staining intensity consistent with the UV absorbance and activity data.

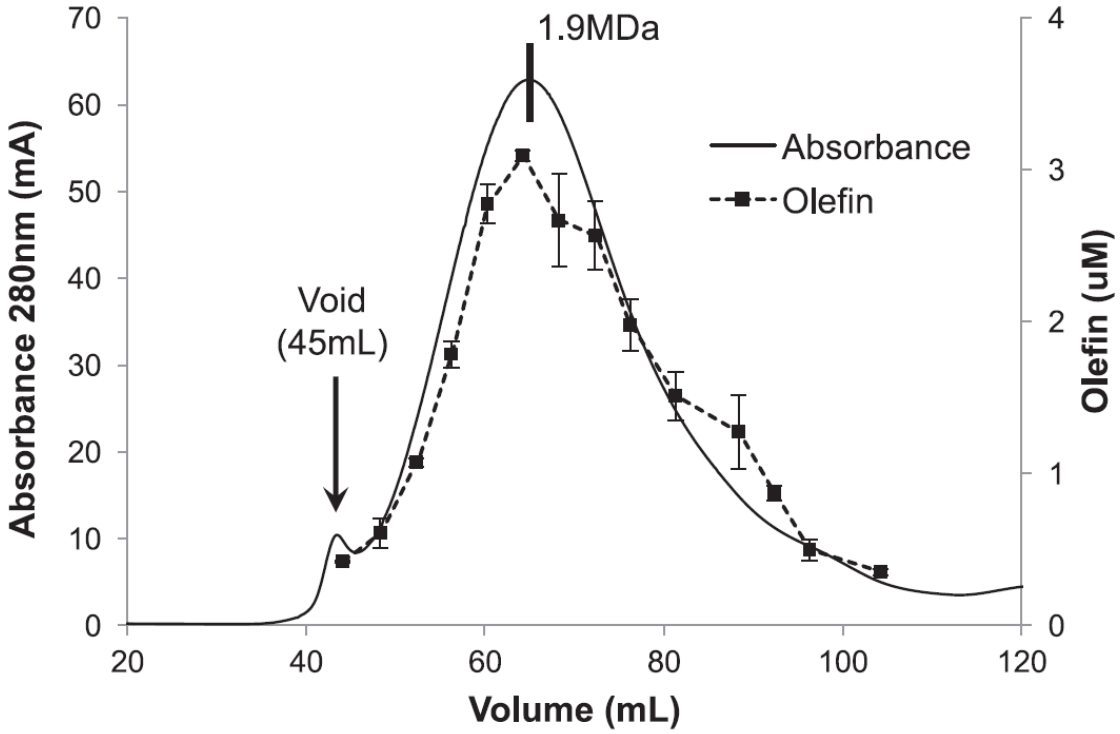


Figure 47: Size exclusion chromatography of Ni^{2+} column-purified OleBCD (tag arrangement 4). The peak absorbance at 65.0 ml corresponded to an estimated size of 1.9 MDa when extrapolated from protein standards. Two hundred microliters of each 4-ml fraction was tested for OleBCD activity by the addition of OleA, myristoyl-CoA, and cofactors.

Electron microscopy imaging of OleBCD assemblies.

The suggested size based on gel filtration prompted analysis of OleBCD by transmission electron microscopy (TEM) using negative staining. This required purification of OleBCD (tag arrangement 4) without glycerol using Ni^{2+}, anti-FLAG, and gel filtration columns to obtain a highly purified sample. Glycerol increases the purification yield of OleBCD assemblies but prevents sample dehydration for effective TEM imaging. Multiple purifications of OleBCD produced very similar electron micrographs, one of which is shown in Figure 48.

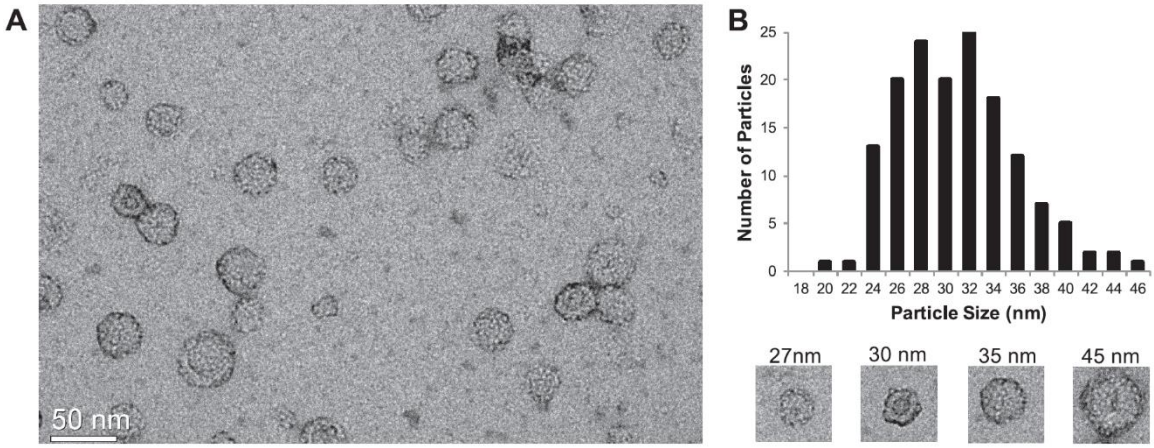


Figure 48: Electron micrograph of OleBCD assemblies with particle size analysis. **A)** OleBCD assemblies by TEM. **B)** Histogram analysis of >150 particles from four different frames, with representative micrographs shown below. Measurement was of the greatest particle diameter. Average assembly size is 30.3 nm, and the median is 29.6 nm.

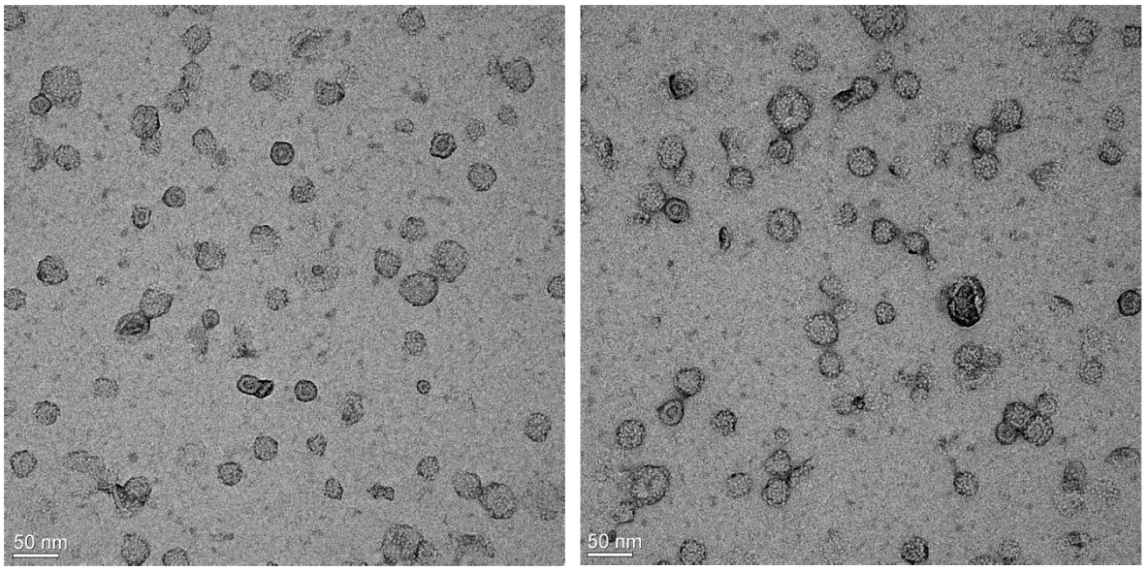


Figure 49: Additional TEM micrographs showing OleBCD (tag arrangement 4) purified over nickel, gel filtration, and anti-FLAG columns.

TEM analysis of the negatively stained OleBCD revealed distinct assemblies for OleBCD, primarily ranging in size from 24 to 40 nm in diameter. The annular hexagonal assemblies with an average diameter of ~27 nm are

particularly striking for their organized structure (Figure 48 & Figure 49). Some of the sample heterogeneity may be different orientations of the same structure on the TEM grid, different size complexes, assembly intermediates, or the result of the dehydration and negative-staining protocol required for TEM. We note that size exclusion chromatography (SEC) does show that OleBCD activity is associated with a broad peak (Figure 47). The relationship between the assemblies pictured in Figure 48 remains unclear, but further work is under way to obtain more high-resolution structural information by cryo-electron microscopic (cryo-EM) methods.

Estimated stoichiometry of the OleBCD assemblies.

To estimate the stoichiometry of each of the OleB, OleC, and OleD proteins within the combined assemblies, gel band intensities of the purified OleBCD, shown in Figure 44C, were compared to band intensities of a gel standard curve made with each purified protein. Figure 50 shows equimolar amounts of separately purified and quantified OleB, OleC, and OleD by SDS-PAGE, as determined by Bradford assay and UV_{280}. Using ImageJ, the intensities of the bands were plotted against protein concentration to generate standard curves for each of the Ole proteins (Figure 51). Comparing the band intensities of the copurified OleBCD assemblies (Figure 44C) with these standard curves revealed a stoichiometry of nearly 4:1:4 of OleB:OleC:OleD within the combined assemblies. Analysis of two other gels gave similar results. Taking this result with the ~1.9 MDa size as assigned by SEC, this would be most consistent

with a single complex composed of 24, 6, and 24 subunits of OleB, OleC, and OleD, respectively. However, the broadness of the SEC peak and the various sizes of the particles observed by cryo-EM might also be consistent with other multiples of the 4:1:4 stoichiometry. It is also possible that a mixture of different component combinations (perhaps assembly intermediates) led to the overall observed 4:1:4 stoichiometry. However, olefin production upon the addition of OleA, cofactors, and substrate occurs in all fractions following SEC, and activity tracks with the amount of protein in the fraction (Figure 47). This indicates that OleB, OleC, and OleD are present in all fractions, although they may not be in the same proportions. We note that the observed correlation between activity and total protein indicates only that the rate-determining step of olefin biosynthesis correlates with total protein.

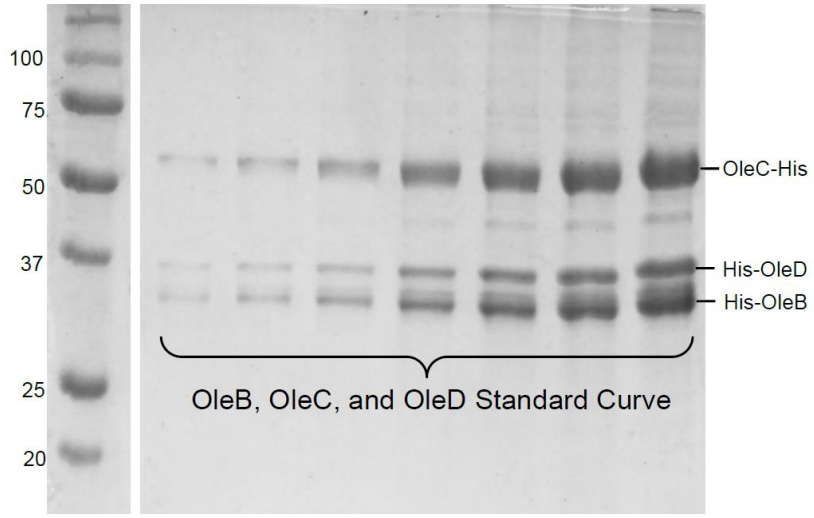
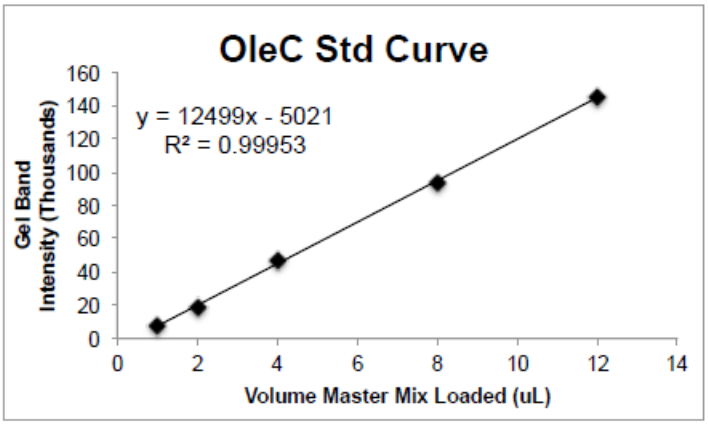
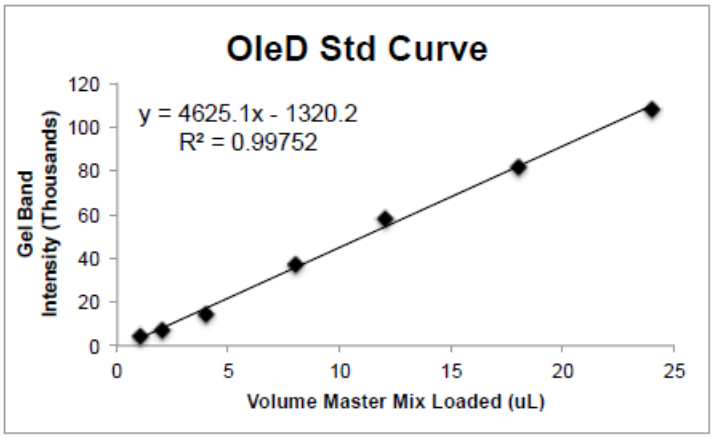


Figure 50: SDS-PAGE standard curve of Ole proteins to determine band staining densities. Separately-expressed OleB, OleC, and OleD were purified over a nickel column. A master mix containing 4 μM of each Ole protein was loaded onto an SDS-PAGE to create this gel standard curve (1 μl, 2 μl, 4 μl, 8 μl, 12 μl, 18 μl, 24 μl). The stoichiometry of the OleBCD combined assemblies was

estimated by comparing band intensities to the generated standard curves (Figure 51).





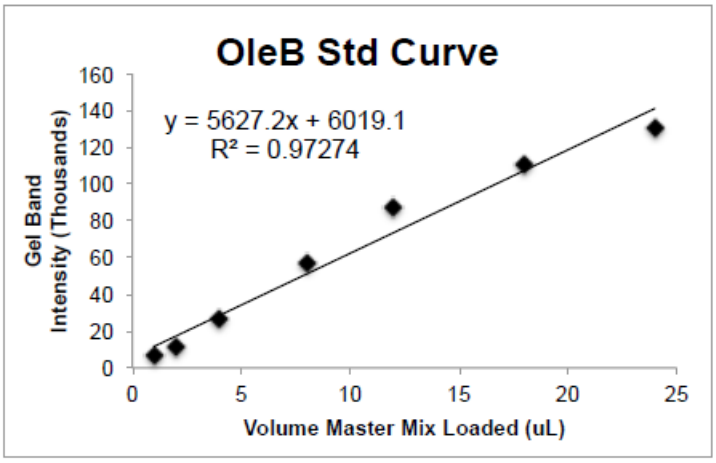


Figure 51: Densitometry analysis of Figure 50. ImageJ was used for the analysis of gel band intensity. Band intensities of the OleBCD complex in Figure 44C were compared to these standard curves to determine a stoichiometric ratio of 4:1:4 OleB:OleC:OleD for the OleBCD combined assemblies.

Discussion

Many biosynthetic pathways are assembled as a single large protein with multiple domains, for example, polyketide synthases and type I fatty acid synthases (97,98). In other cases, individually expressed proteins self-assemble in the cytosol, for example, the lumazine synthase-riboflavin synthase protein complex (99,100). The structure and function of a number of polyketide synthase and fatty acid synthase systems have been studied in significant detail (101-103). In contrast, the higher-order structure of hydrocarbon biosynthetic machinery has been far less characterized.

In the present study, the OleABCD enzymes from *X. campestris* were expressed and purified individually and shown to reconstitute long-chain olefin biosynthesis. Coexpression studies demonstrated that OleBCD forms active stable assemblies. While hundreds of bacteria are thought to produce long-chain head-to-head hydrocarbons, *X. campestris* has been used as a model system because the enzymes are relatively stable and have been shown to possess broad substrate specificity, making them suitable for bioengineering (13). OleA purifies as a stable soluble dimer, and the structure and mechanism of this component have been the subject of several studies (15,25,54). OleA catalyzes the Claisen condensation of fatty acyl-CoA substrates, producing a β-keto acid that feeds into the OleBCD enzymes (15).

There are several reasons that multienzyme OleBCD assemblies might be functionally superior to individual OleB, OleC, and OleD proteins operating independently within a bacterial cell. First, a multienzyme assembly of Ole proteins could protect the cell by sequestering the reactive Ole pathway β-lactone intermediate and preventing a nonspecific reaction with cytosolic proteins. The β-lactone moiety is common in microbial natural products and can function as an antibiotic by reacting with active-site residues in essential esterase and protease enzymes (82,86). The β-lactone tetrahydrolipstatin from *Streptomyces toxytricini* is very similar in structure to the β-lactone produced by OleC, and its biosynthesis is encoded by an *ole*-like gene cluster that lacks the *oleB* decarboxylase gene (50,82). Tetrahydrolipstatin is a potent human pancreatic lipase inhibitor and is the only FDA-approved over-the-counter antiobesity drug. Since the β-lactone of olefin biosynthesis is only an intermediate, the sequestration of a potentially harmful compound by an enzyme complex likely affords sufficient protection to the bacterium. It is also relevant that *oleB* and *oleC* are a gene fusion in all actinobacteria, which comprise approximately 30% of known *ole*-gene-containing organisms. It is unknown if assemblies of OleBCD form in these bacteria, as the enzyme fusion should allow the direct transfer of the β-lactone from the OleC to OleB domains for decarboxylation. Fusion creates a 1:1 stoichiometry between OleB and OleC, but our protein quantitation for the *X. campestris* multienzyme assemblies suggests a

4:1 relationship, respectively, suggesting there may be different configurations between species.

An additional benefit of an OleBCD complex would be to promote efficient trafficking of the highly hydrophobic intermediates that comprise the Ole pathway from one active site to another. It is common for the multistep biosynthesis or degradation of hydrophobic compounds to be accomplished by either membrane-bound enzymes or enzyme complexes (97,98). However, sequence analysis of Ole proteins did not predict any transmembrane domains or lipid anchor sequences. Moreover, the enzymes have been observed in the soluble protein fraction of the native host *X. campestris* and in the recombinant host *E. coli*. An example of a well-characterized bacterial complex producing and transferring hydrophobic intermediates is the fatty acid synthesis (fatty acid synthase type I [FAS I]) machinery from *Mycobacterium tuberculosis* (104). Cryo-EM studies of FAS I from *M. tuberculosis* revealed a 2.0-MDa homohexameric enzyme complex with a diameter of 25 nm and the same hexagonal symmetry observed in our annular assemblies by TEM of the OleBCD assemblies. However, the FAS I complex uses an acyl carrier protein within a central cavity to ferry enzyme-tethered intermediates between active sites for iterative fatty acid elongation reactions. While a central cavity or porous channeled structure is possible based on the size and molecular weight of the OleBCD assemblies, there is no evidence for any enzyme-tethered intermediate shuttle in the Ole biosynthetic pathway at this time.

In summary, we propose the following model for long-chain olefin biosynthesis in *X. campestris* (Figure 52). OleA receives fatty acyl-CoA products from the fatty acid biosynthetic machinery, acting as a shuttle between that complex and an OleBCD complex. It is likely that OleA has some binding affinity for an OleBCD complex, as suggested by assemblies of OleBCD showing weaker affinity for His_6-OleA on the Ni^{2+} column. However, this interaction is not sufficiently strong to survive passage through several chromatographic steps. The current data support the idea that the consecutively catalyzed OleD-OleC-OleB reactions occur sequestered within multienzyme structures.

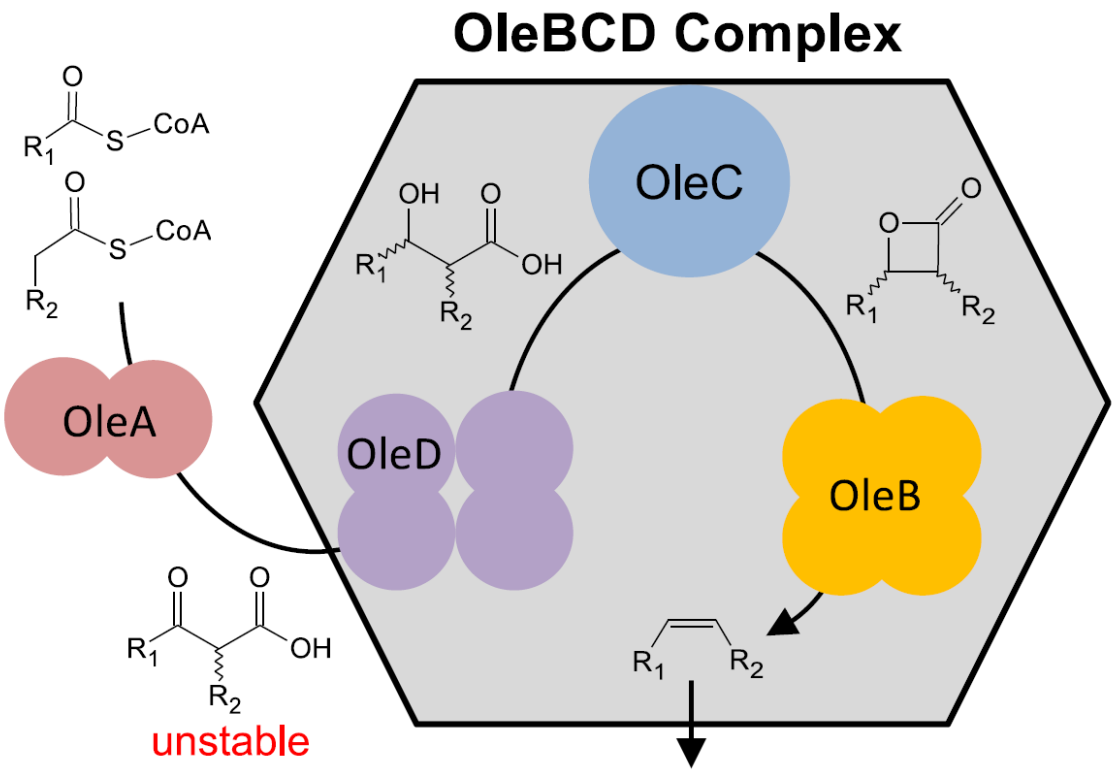


Figure 52: Model for the biosynthesis of long-chain head-to-head olefins in *X. campestris* showing that OleA acts predominantly as a soluble cytosolic dimer, condensing acyl-CoAs, docking, and providing its reaction product to the OleBCD complex, which ultimately releases a long-chain olefin.

The product olefin is ultimately found in the membrane fraction, and the mechanism of transfer from the soluble OleBCD assemblies to the membrane remains to be elucidated. Further studies by cryo-EM will help reveal more details of the OleBCD assembly architecture and the relationship between the component enzymes. Longchain hydrocarbon production has been confirmed experimentally in at least 20 different bacteria (12,13,15,16,88,96), and sequence analysis reveals *oleABCD* gene clusters in hundreds of other bacteria, suggesting that multienzyme OleBCD structures may be widespread among microorganisms.

Materials and methods

Chemicals

All reagents were purchased from Sigma-Aldrich, St. Louis, MO, with the following exceptions. NADPH was purchased from EMD Millipore, Billerica, MA. Glycerol was purchased from IBI Scientific, Peosta, IA. Synthetic *cis*- and *trans*-olefins were prepared as described previously (50).

Cloning, expression, and purification of Ole proteins

All protein sequences for *ole* genes used in this study were from *X. campestris* pv. *campestris* ATCC 33913 (OleA, accession no. NP_635607.1; OleB, accession no. NP_635611.1; OleC, accession no. NP_635613.2; and OleD, accession no. NP_635614.1). *E. coli* codon-optimized genes were ordered

from DNA2.0, Inc. Polymerase, restriction enzymes, ligase, and supercompetent cells were purchased from New England BioLabs and used according to the manufacturer's instructions. The four vector backbones, pET28b$^+$, pET30b$^+$, pCOLA, and pCDF, were obtained from Novogen. A complete list of constructs, gene insertion sites, and antibiotic selectable markers from this study can be found in Table 15.

Table 15: Constructs, insertion sites, and selectable markers used in this study.

Plasmids	Description	Source
pCOLA Duet (kan)		
oleA/oleB	Inducible expression of a N-terminal His_{6x} tagged *oleA* with *oleB*. SacI:PstI (*oleA*)/BglII:XhoI (*oleB*)	This study
oleB/oleA	Inducible expression of a N-terminal His_{6x} tagged *oleB* with *oleA*. BamHI:SacI (*oleB*)/NceI:XhoI (*oleA*)	This study
pCDF Duet (sm)		
oleD/oleC	Inducible expression of *oleD* with *oleC*. NcoI:BamHI (*oleD*)/NdeI:XhoI (*oleC*)	This study
oleD/oleC	Inducible expression of *oleD* with *oleC* with a C-terminal FLAG tag. NcoI:BamHI (*oleD*)/NdeI:XhoI (*oleC*)	This study
oleB/oleC	Inducible expression of *oleB* with *oleC* with a C-terminal FLAG tag. NcoI:SacI (*oleB*)/NdeI:XhoI (*oleC*)	This study
pET28b+ (kan)		
oleA	Inducible expression of *oleA* a N-terminal His_{6x} tag. Cloned into NdeI:XhoI site	(1)
oleB	Inducible expression of *oleB* a N-terminal His_{6x} tag. Cloned into NdeI:BamHI site	This study
oleD	Inducible expression of *oleD* a N-terminal His_{6x} tag. Cloned into NdeI:XhoI site	(1)
pET30b+ (kan)		
oleC	Inducible expression of *oleC* a C-terminal His_{6x} tag. Cloned into NdeI:HindIII site	(50)
oleD	Inducible expression of *oleD* a C-terminal His_{6x} tag. Cloned into NdeI:XhoI site	This study

kanamycin (kan), streptomycin (sm)

The following general expression and purification protocols were developed for all Ole enzyme configurations. Plasmids were transformed into *E. coli* BL21(DE3) cells (Invitrogen) and plated on Luria-Bertani (LB) agar with appropriate antibiotic selection (50 µg/µl). Single colonies were selected and grown at 37°C in 5 ml of LB overnight as starter cultures. LB medium (1.0 liter) containing antibiotic was inoculated with a 5-ml starter culture and grown at 37°C until an optical density at 600 nm (OD_{600}) of 0.50 was reached. Protein expression was then induced with 100 µM isopropyl β-D-1-thiogalactopyranoside (IPTG), and growth was shifted to 16°C overnight. Cells were harvested by centrifugation and resuspended in 5 ml of buffer containing 200 mM NaCl, 20 mM HNa_2PO_4 (pH 7.4), and 10% glycerol (buffer) per liter of LB medium. Sonication or French pressure cell lysis was used to disrupt the cells, followed by centrifugation at 35,000 x *g*. The supernatant liquid was filtered through 0.45-µm- and 0.22-µm-pore size filters (Millipore) and loaded onto a GE HisTrap HP 5-ml column (His column). Gradient elution in buffer with 500 mM imidazole (pH 7.4) was achieved using an ÄKTA fast protein liquid chromatography instrument (General Electric Company). Ole proteins typically eluted around 200 mM imidazole. Coexpressed protein mixtures containing a DYKDDDDK (FLAG) tag were then manually loaded onto a column containing 2.5 ml of anti-FLAG M2 resin (Sigma-Aldrich) (anti-FLAG column). The anti-FLAG column was washed with 8 column volumes of buffer before eluting with 10 ml of 100 µg/ml FLAG peptide. The protein in the solution was concentrated using 30,000-molecular-

weight-cutoff (30,000-MWCO) spin filters (Amicon) and frozen at -80°C for later use.

The following protein-specific modifications were made to the above-mentioned protocol. Whenever buffer additives were required, as described below, they were included for all purification steps. For OleBCD assemblies, vector combinations, shown in Figure 44A, were transformed into *E. coli* under both kanamycin and streptomycin selection. No detergents were needed during purification of Ole assemblies, but 10% (vol/vol) glycerol was found to increase protein yield. For electron microscopy experiments, glycerol was omitted from the purification buffer. For OleA, after IPTG induction, cells containing pET28b$^+$ *oleA* with an N-terminal His_6 tag were grown at 37°C for 4 hr before purification in buffer. For OleB, buffer required 0.05% (wt/vol) Triton X-100 (critical micelle concentration [cmc], ~0.02% [wt/vol]) for purification of pET28b$^+$ expressed OleB with a His_6 N-terminal tag. Triton X-100 interfered with A_{280} measurements, necessitating that protein elution be followed by the Bradford assay and protein gels (105). Protein could not be concentrated without precipitation. For OleC, an N-terminal His_6 tag on OleC does not purify, necessitating the C-terminal His_6 tag of pET30b$^+$ for expression and purification. For OleD, buffer containing 0.02% (wt/vol) Tween 20 (cmc, ~0.007% [wt/vol]) was required to purify OleD with an N-terminal His_6 tag expressed from a pET28b$^+$ vector. Protein could not be concentrated without precipitation.

Olefin activity assays

Ole enzyme activity assays were conducted in glass vials at room temperature with 500 µl of buffer (200 mM NaCl, 20 mM $NaPO_4$ pH 7.4). Reactions were conducted in duplicate and contained 10 µg of OleA and 30 µg of copurified OleBCD, as measured by the Bradford assay. Assays involving separately purified OleA, OleB, OleC, and OleD contained 10 µg of OleA, 12.1 µg of OleB, 5.1 µg of OleC, and 12.8 µg of OleD to mimic the putative 4:1:4 B:C:D stoichiometry of 30 µg of copurified OleBCD. ATP, NADPH, and $MgCl_2$ were added to final concentrations of 1 mM, 500 µM, and 1 mM, respectively, in all assays. The slight difference in buffer additives to Ole proteins was ignored, as they were found not to significantly affect reactions. Reactions were initiated by the addition of myristoyl-CoA to a concentration of 50 µM and quenched with 500 µl of ethyl acetate at appropriate time points. Five microliters of organic phase was removed and analyzed by GC, using both a flame ionization detector (FID) and a mass spectrometry (MS) detector, as previously described (50). The program conditions were as follows: start temperature, 100°C; ramp rate, 10°C/min; final temperature, 320°C; hold time, 5 min; and total time, 27 min. A synthetic olefin standard, *cis*-13-heptacosene, was synthesized as described previously (50). The retention time of *cis*-13-heptacosene was 18.26 min.

Mass spectrometry protein identification

Ole proteins were separated using 12.5% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Bands were excised, trypsin

digested, and submitted to the University of Minnesota Center for Mass Spectrometry and Proteomics for peptide identification using liquid chromatography-mass spectrometry (106). Peptides were initially separated by a Paradigm Platinum Peptide Nanotrap (Michrom Bioresources, Inc.) precolumn, followed by an analytical capillary column (100 µm by 12 cm) packed with C_{18} resin (5 µm, 200 Å MagicC18AG; Michrom Bioresources, Inc.). Mass spectrometry was performed on an LTQ (Thermo Electron Corp., San Jose, CA).

Immunoblots for OleC in native X. campestris lysate

Polyclonal antibodies against an *X. campestris* OleC peptide (AIDDAAIPEWSGVR) were raised by GenScript in rabbit. A secondary horseradish peroxidase-conjugated donkey anti-rabbit IgG antibody was purchased from Jackson ImmunoResearch Laboratories. *X. campestris* (ATCC 33913) for immunoblots was grown in 250 ml of LB medium at 30°C before harvesting and sonication generated the *X. campestris* lysate. OleC and OleBCD (Figure 44A, tag arrangement 4) were purified as described above. Native gels were made as described by Nadano et al. (107) and contained 1.9% polyacrylamide and 0.75% agarose in 40 mM Tris-acetate buffer at pH 8.4. Protein was mixed with loading buffer (15% sucrose in Tris-acetate) and run at 100 V. Gels were transferred to a nitrocellulose membrane and washed thoroughly with 200 mM NaCl, 20 mM $NaPO_4$ (pH 7.4), and 0.1% Tween 20 as a blocking agent. Membranes were incubated with the primary antibody overnight, washed, and incubated with the secondary antibody for 1 h. ECL Plus Western

blotting substrate (Pierce) and Classic Blue autoradiography film BX (MidSci) were used to visualize the membrane.

Gel filtration chromatography

A HiLoad 16/600 Superdex 200 (GE; normal column) and a HiPrep 16/60 Sephacryl S-400 HR (GE; ultrahigh-molecular-weight [ultrahigh-MW] column) were used to separate the proteins and determine molecular weight. The normal small-size column can separate globular proteins up to 600 kDa, while the ultrahigh-MW column can separate globular proteins up to 8.0 MDa. A gel filtration high-MW (HMW) calibration kit was purchased from GE to calibrate the gel filtration columns (up to a molecular mass of 669 kDa), with higher molecular masses being extrapolated. Typically, 2.0 mg of protein was loaded onto columns and run according to the manufacturer's specifications.

Electron microscopy of OleBCD assemblies

Following His affinity column, anti-FLAG, and gel filtration purification, OleBCD assemblies were characterized using transmission electron microscopy (TEM). Grids (400-mesh Cu-carbon grids) were glow discharged at 15 mA and 39 Pa for 1 min using a Pelco easiGlow glow discharge cleaning system (Ted Pella, Inc., Redding, CA). Samples containing 0.08 mg/ml OleBCD were drop-cast onto these grids and blotted with filter paper to remove excess sample. Grids were washed twice with deionized water and blotted dry. Uranyl formate (0.05%) stain was added to each grid and blotted a final time. Negative-stain micrographs were collected using a field emission gun (FEG) Tecnai G2 F30

TEM (FEI, Hillsboro, OR). Images were recorded at room temperature on a Gatan 4k by 4k charge-coupled-device (CCD) camera (Gatan, Inc., Pleasanton, CA). ImageJ was used to analyze the sizes of >150 particles in four TEM images. The size was recorded as the greatest diameter of the particle.

SDS-PAGE standard curve

Quantification of staining intensities of the Ole proteins on SDS-PAGE was carried out using individually purified Ole proteins, as described above. Concentrations of OleB, OleC, and OleD were determined by the Bradford assay and supported by UV absorbance at 280 nm (A_{280}). OleB required Triton X-100 to prevent protein precipitation and so could not be verified by A_{280}. A 4.0 µM master mix of the three proteins was created and loaded in different volumes into wells of a 12.5% SDS-PAGE gel. Gels were stained with SimplyBlue SafeStain (Life Technologies), and images were analyzed using ImageJ.

CHAPTER 8

OleBCD conclusions

Our work on the remaining enzymes in the olefin biosynthesis pathway has revealed novel features of this enzymes pathway. New enzyme classifications and large multienzyme assemblies have led to a deeper understanding of the mechanisms and interactions between OleB, OleC, and OleD. These investigations have provided the basic knowledge that may aid in the discovery of similar biosynthesis pathways in related organisms and the potential application of using OleC and newly identified enzymes to produce natural product therapeutics. The work on OleBCD also opens the door for future structural understanding of complex formation and hydrophobic substrate trafficking between enzyme active sites.

Conclusions from OleC β-lactone investigation

Thorough analysis and characterization of intermediates and products reveal that OleC produces long-chain β-lactones. Due to the thermal lability of β-lactones, anaylsis of the OleC product by GC/MS resulted in artefactual olefins. The use of synthetic standards and ^{1}H NMR demonstrate that *syn*- and *anti*-β-lactones are produced by OleC from *syn*- and *anti*-β-hydroxy acids, respectively. This represents the first described β-lactone synthetase. Bioinformatic analysis of

known β-lactone-producing gene clusters reveal enzymes homologous to OleC, suggesting a common mechanism of β-lactone synthetase activity. The role of OleB in catalyzing the final decarboxylation of β-lactones to form olefins is also proposed, representing a second novel enzyme function within the pathway.

Conclusions from OleBCD assembly investigation

All four enzymes in the olefin biosynthesis pathway are shown to be necessary for olefin production. The co-expression and purification of OleBCD results in the isolation of multienzyme assemblies that produce olefin with addition of separately-purified OleA. By negative stain electron microscopy, OleBCD assemblies show distinct, large structures that vary in size, consistent with molecular weights predicted by size exclusion chromatography. Support for the native formation of these assemblies may be to protect highly reduced carbon intermediates during catalysis, preserve the stereochemical configuration of metabolites, and sequester the highly reactive β-lactone produced by OleC.

Future directions

A crystal structure of OleC would complement its newly discovered function as a β-lactone synthetase because it would represent the first structure of a new enzyme class. Crystallographic efforts are ongoing. Past attempts to crystallize OleC from *Xanthomonas campestris, Stentrophomonas maltophilia,* and the natural fusion of OleB/C from *Micrococcus luteus* (full-length and

truncations) have been unsuccessful or irreproducible (18). Co-crystallization trials utilizing β-hydroxy acids, β-lactones, ATP, and AMP have also failed to produce well-diffracting crystals. The only dataset collected to date is that for *S. maltohphilia* OleC at 3.4 Å resolution, and the crystallization was not reproducible (18). My attempts to phase these data using molecular replacement have not been successful despite newly published ANL superfamily structures and updates to crystallographic software. This demonstrates that the structure of OleC will need to be solved using experimental phasing. Attempts to purify selenomethionine-labeled OleC are currently underway. Wide screening of crystallization conditions for Se-labeled OleC from multiple species will need to be performed to identify the ideal conditions for growing well diffracting protein crystals. Structures of OleC in the presence of substrates, intermediates, and products will be useful in elucidating the mechanism of β-lactone formation. An intriguing experiment would be to grow co-crystals of OleC with β-hydroxy acids. These co-crystals could then be soaked with ATP prior to flash freezing and data collection. This soak experiment would allow for partial turnover of β-hydroxy acid to β-lactone. Trapped intermediates, such as adenylated β-hydroxy acid, would yield important information about which amino acid residues are important for β-lactone ring closure and how the alkyl chains of the substrate are bound within the enzyme. It would also directly demonstrate the site of adenylation, which is hypothesized as the carboxylic acid moiety of the β-hydroxy acid (50).

An overall structure of the OleBCD assembly is another goal of this project. As outlined in Chapter 7, negative stain transmission electron microscopy was used to visualize the assemblies. The heterogeneity of the sample and large assembly sizes of around 2 MDa mean that X-ray crystallography is likely not a viable option to solve its structure. Instead, cryo-electron microscopy (cryo-EM) is being employed to solve a relatively high resolution structure of OleBCD. Ongoing work is being conducted to improve the monodispersity and size population homogeneity of purified OleBCD. Expression of OleBCD from *M. luteus* is also being explored because its natural OleB/C fusion maintains a 1:1 stoichiometry between OleB and OleC. Sample grid quality is being assessed by negative-stain EM in the Characterization Facility at the University of Minnesota under the mentorship of Dr. Wei Zhang. The goal is to transition our sample preparation into one viable for cryo-EM analysis. Images can then be collected using the Titan Krios G2 microscope at the Hormel Institute in Austin, MN. Three-dimensional reconstruction from averaged two-dimensional micrographs of oriented particles will enable solution of the structure of OleBCD.

Another goal involving OleBCD is to determine its subcellular localization. As the pathway produces very hydrophobic hydrocarbons that affect bacterial adaptation to growth in cold temperatures, it is likely that olefins are inserted into the membrane (7). It would follow that the OleBCD would associate near the membrane. In order to accurately describe the location of OleBCD *in vivo*, I have prepared fluorescently labeled Ole proteins for super-resolution fluorescence

microscopy in intact bacterial cells. This project is in collaboration with Dr. Andreas Gahlmann at the University of Virginia. This technique is able to overcome the diffraction limitations of light microscopes and is sensitive to single molecules within a living bacterial cell (108). Dr. Gahlmann's instrument is able to localize single fluorophores within a bacterial cell and calculate their diffusion coefficients by tracking their motion over a period of time. Applying this to the olefin biosynthesis pathway, I have constructed plasmids that express fluorescent protein-tagged OleA, OleC, and OleD from *Shewanella oneidensis* MR-1. These tagged proteins were transformed into specific gene-deletion strains of *S. oneidensis* (e.g. eYFP-OleA expressed in *ΔoleA* background) (Figure 53). Deletion strains for *oleA* and *oleC* have been generated, but the *ΔoleD* strain needs to be prepared (7). The subcellular localization and diffusion coefficient of the OleBCD assembly will then be analyzed by Dr. Gahlmann's group. Complementation studies show that fluorescent protein tags do not interrupt olefin production; however, the effect these fluorescent protein tags have on formation of OleBCD assemblies needs to be investigated. To do this, we will overexpress fluorescent protein-tagged *S. oneidensis* OleBCD in *E. coli* for EM analysis. The same will be done with *X. campestris* OleBCD for comparison.

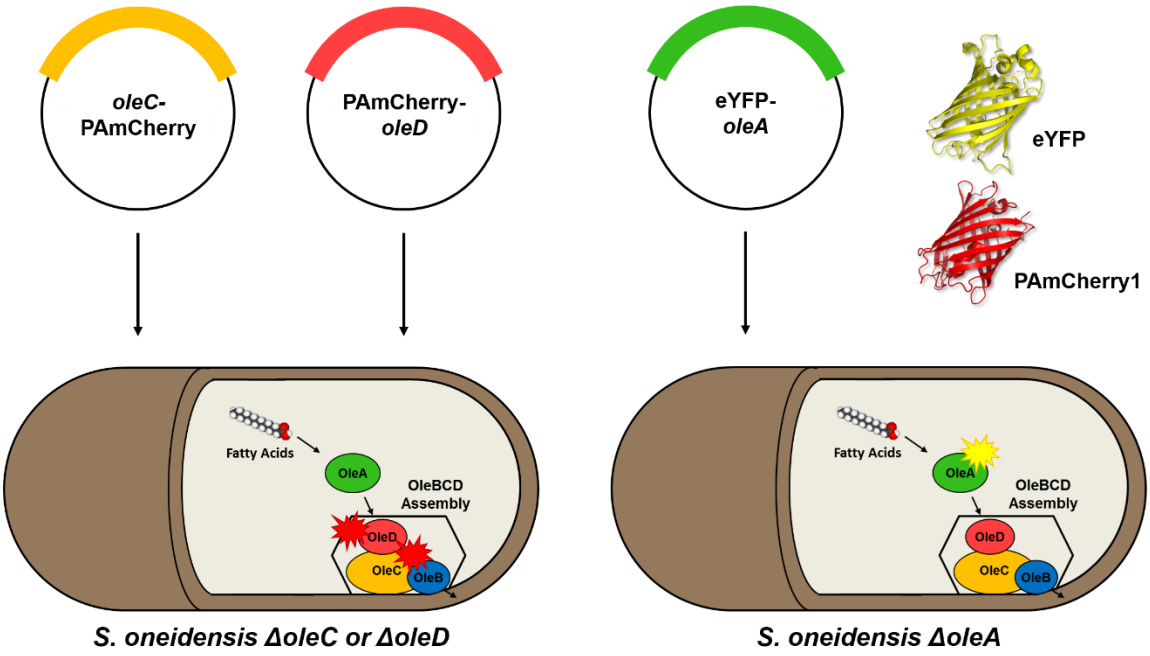


Figure 53: Outline of *Shewanella oneidensis* MR-1 super-resolution fluorescence microscopy localization experiment. Complementation of specific Ole protein knockout strains with plasmids expressing fluorescent protein-tagged OleA, OleC, and OleD will demonstrate the subcellular localization of OleBCD (assembly shown as black hexagon). eYFP and PAmCherry1 fluorescent protein models are from PDB 3V3D and 3KCT, respectively.

References

1. Yergin, D. (2011) *The quest : energy, security and the remaking of the modern world*, Penguin Press, New York
2. Lee, S. K., Chou, H., Ham, T. S., Lee, T. S., and Keasling, J. D. (2008) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* **19**, 556-563
3. Hill, J., Nelson, E., Tilman, D., Polasky, S., and Tiffany, D. (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11206-11210
4. Rude, M. A., and Schirmer, A. (2009) New microbial fuels: a biotech perspective. *Curr. Opin. Microbiol.* **12**, 274-281
5. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**, 804-807
6. Notteboom, T., Carriou, P. (2009) Fuel surcharge practices of container shipping lines: Is it about cost recovery or revenue making? *Proc. IAME* **2009**
7. Sukovich, D. J., Seffernick, J. L., Richman, J. E., Hunt, K. A., Gralnick, J. A., and Wackett, L. P. (2010) Structure, function, and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1. *Appl. Environ. Microbiol.* **76**, 3842-3849
8. Albro, P. W., and Dittmer, J. C. (1969) Biochemistry of long-chain, nonisoprenoid hydrocarbons. III. Metabolic relation of long-chain fatty acids and hydrocarbons and other aspects of hydrocarbon metabolism in *Sarcina lutea*. *Biochemistry* **8**, 1913-1918
9. Friedman, L., and Da Costa, B. (2010) Hydrocarbon-producing genes and methods of their use. WO Patent 2008147781
10. Friedman, L., and Rude, M. (2008) Process for producing low molecular weight hydrocarbons from renewable resources. WO Patent 2008113041
11. Frias, J. A., Richman, J. E., and Wackett, L. P. (2009) C29 olefinic hydrocarbons biosynthesized by *Arthrobacter* species. *Appl. Environ. Microbiol.* **75**, 1774-1777
12. Beller, H. R., Goh, E. B., and Keasling, J. D. (2010) Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Appl. Environ. Microbiol.* **76**, 1212-1223
13. Sukovich, D. J., Seffernick, J. L., Richman, J. E., Gralnick, J. A., and Wackett, L. P. (2010) Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Appl. Environ. Microbiol.* **76**, 3850-3862
14. Haapalainen, A. M., Merilainen, G., and Wierenga, R. K. (2006) The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends Biochem. Sci.* **31**, 64-71
15. Frias, J. A., Richman, J. E., Erickson, J. S., and Wackett, L. P. (2011) Purification and characterization of OleA from *Xanthomonas campestris*

and demonstration of a non-decarboxylative Claisen condensation reaction. *J. Biol. Chem.* **286**, 10930-10938
16. Bonnett, S. A., Papireddy, K., Higgins, S., del Cardayre, S., and Reynolds, K. A. (2011) Functional characterization of an NADPH dependent 2-alkyl-3-ketoalkanoic acid reductase involved in olefin biosynthesis in *Stenotrophomonas maltophilia*. *Biochemistry* **50**, 9633-9640
17. Gulick, A. M. (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811-827
18. Frias, J. A., Goblirsch, B. R., Wackett, L. P., and Wilmot, C. M. (2010) Cloning, purification, crystallization and preliminary X-ray diffraction of the OleC protein from *Stenotrophomonas maltophilia* involved in head-to-head hydrocarbon biosynthesis. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1108-1110
19. Kancharla, P., Bonnett, S. A., and Reynolds, K. A. (2016) *Stenotrophomonas maltophilia* OleC-Catalyzed ATP-Dependent Formation of Long-Chain Z-Olefins from 2-Alkyl-3-hydroxyalkanoic Acids. *ChemBioChem* **17**, 1426-1429
20. Christenson, J. K., Robinson, S. L., Engel, T. A., Richman, J. E., Kim, A. N., and Wackett, L. P. (2017) OleB from bacterial hydrocarbon biosynthesis is a β-lactone decarboxylase sharing key features with haloalkane dehalogenases. *Biochemistry*
21. Musayev, F., Sachdeva, S., Scarsdale, J. N., Reynolds, K. A., and Wright, H. T. (2005) Crystal structure of a substrate complex of *Mycobacterium tuberculosis* β-ketoacyl-acyl carrier protein synthase III (FabH) with lauroyl-coenzyme A. *J. Mol. Biol.* **346**, 1313-1321
22. Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **6**, 775-784
23. Theisen, M. J., Misra, I., Saadat, D., Campobasso, N., Miziorko, H. M., and Harrison, D. H. (2004) 3-hydroxy-3-methylglutaryl-CoA synthase intermediate complex observed in "real-time". *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16442-16447
24. Modis, Y., and Wierenga, R. K. (2000) Crystallographic analysis of the reaction pathway of *Zoogloea ramigera* biosynthetic thiolase. *J. Mol. Biol.* **297**, 1171-1182
25. Goblirsch, B. R., Frias, J. A., Wackett, L. P., and Wilmot, C. M. (2012) Crystal structures of *Xanthomonas campestris* OleA reveal features that promote head-to-head condensation of two long-chain fatty acids. *Biochemistry* **51**, 4138-4146
26. Kresovic, D., Schempp, F., Cheikh-Ali, Z., and Bode, H. B. (2015) A novel and widespread class of ketosynthase is responsible for the head-to-head condensation of two acyl moieties in bacterial pyrone biosynthesis. *Beilstein J. Org. Chem.* **11**, 1412-1417

27. Sachdeva, S., Musayev, F. N., Alhamadsheh, M. M., Scarsdale, J. N., Wright, H. T., and Reynolds, K. A. (2008) Separate entrance and exit portals for ligand traffic in *Mycobacterium tuberculosis* FabH. *Chem. Biol.* **15**, 402-412
28. Campobasso, N., Patel, M., Wilding, I. E., Kallender, H., Rosenberg, M., and Gwynn, M. N. (2004) *Staphylococcus aureus* 3-hydroxy-3-methylglutaryl-CoA synthase: crystal structure and mechanism. *J. Biol. Chem.* **279**, 44883-44888
29. Wackett, L. P., and Wilmot, C.M. (2015) Hydrocarbon biosynthesis in microorganisms. in *Direct Microbial Conversion of Biomass to Advanced Biofuels* (Himmel, M. E. ed.), Elsevier, Amsterdam, The Netherlands. pp 442-470
30. Heath, R. J., and Rock, C. O. (2002) The Claisen condensation in biology. *Nat. Prod. Rep.* **19**, 581-596
31. Wang, J., Soisson, S. M., Young, K., Shoop, W., Kodali, S., Galgoci, A., Painter, R., Parthasarathy, G., Tang, Y. S., Cummings, R., Ha, S., Dorso, K., Motyl, M., Jayasuriya, H., Ondeyka, J., Herath, K., Zhang, C., Hernandez, L., Allocco, J., Basilio, A., Tormo, J. R., Genilloud, O., Vicente, F., Pelaez, F., Colwell, L., Lee, S. H., Michael, B., Felcetto, T., Gill, C., Silver, L. L., Hermes, J. D., Bartizal, K., Barrett, J., Schmatz, D., Becker, J. W., Cully, D., and Singh, S. B. (2006) Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* **441**, 358-361
32. Sridharan, S., Wang, L., Brown, A. K., Dover, L. G., Kremer, L., Besra, G. S., and Sacchettini, J. C. (2007) X-ray crystal structure of *Mycobacterium tuberculosis* β-ketoacyl acyl carrier protein synthase II (mtKasB). *J. Mol. Biol.* **366**, 469-480
33. Pan, H., Tsai, S., Meadows, E. S., Miercke, L. J., Keatinge-Clay, A. T., O'Connell, J., Khosla, C., and Stroud, R. M. (2002) Crystal structure of the priming β-ketosynthase from the R1128 polyketide biosynthetic pathway. *Structure* **10**, 1559-1568
34. Alexson, S. E., and Nedergaard, J. (1988) A novel type of short- and medium-chain acyl-CoA hydrolases in brown adipose tissue mitochondria. *J. Biol. Chem.* **263**, 13564-13571
35. Ellman, G. L. (1958) A colorimetric method for determining low concentrations of mercaptans. *Arch. Biochem. Biophys.* **74**, 443-450
36. Price, A. C., Choi, K. H., Heath, R. J., Li, Z., White, S. W., and Rock, C. O. (2001) Inhibition of β-ketoacyl-acyl carrier protein synthases by thiolactomycin and cerulenin. Structure and mechanism. *J. Biol. Chem.* **276**, 6551-6559
37. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., and Larsen, S. (2001) Structures of β-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **9**, 233-243

38. Zhang, Y. M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O., and White, S. W. (2001) Identification and analysis of the acyl carrier protein (ACP) docking site on β-ketoacyl-ACP synthase III. *J. Biol. Chem.* **276**, 8231-8238
39. Erol, O., Schaberle, T. F., Schmitz, A., Rachid, S., Gurgui, C., El Omari, M., Lohr, F., Kehraus, S., Piel, J., Muller, R., and Konig, G. M. (2010) Biosynthesis of the Myxobacterial Antibiotic Corallopyronin A. *ChemBioChem* **11**, 1253-1265
40. Sucipto, H., Wenzel, S. C., and Muller, R. (2013) Exploring Chemical Diversity of α-Pyrone Antibiotics: Molecular Basis of Myxopyronin Biosynthesis. *ChemBioChem* **14**, 1581-1589
41. Sucipto, H., Sahner, J. H., Prusov, E., Wenzel, S. C., Hartmann, R. W., Koehnke, J., and Muller, R. (2015) In vitro reconstitution of α-pyrone ring formation in myxopyronin biosynthesis. *Chemical Science* **6**, 5076-5085
42. Miziorko, H. M., Shortle, D., and Lane, M. D. (1976) Trapping of a novel coenzyme A containing intermediate of 3-hydroxy-3-methylglutaryl-CoA synthase. *Biochem. Biophys. Res. Commun.* **69**, 92-98
43. Chester, N., and Marshak, D. R. (1993) Dimethyl sulfoxide-mediated primer Tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Anal. Biochem.* **209**, 284-290
44. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221
45. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501
46. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21
47. Schrodinger, LLC. (2015) The PyMOL Molecular Graphics System, Version 1.8.05.
48. Wackett, L. P. (2011) Engineering microbes to produce biofuels. *Curr. Opin. Biotechnol.* **22**, 388-393
49. Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B., and Keasling, J. D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559-562
50. Christenson, J. K., Richman, J. E., Jensen, M. R., Neufeld, J. Y., Wilmot, C. M., and Wackett, L. P. (2017) β-Lactone Synthetase Found in the Olefin Biosynthesis Pathway. *Biochemistry* **56**, 348-351

51. Christenson, J. K., Jensen, M. R., Goblirsch, B. R., Mohamed, F., Zhang, W., Wilmot, C. M., and Wackett, L. P. (2017) Active Multienzyme Assemblies for Long-Chain Olefinic Hydrocarbon Biosynthesis. *J. Bacteriol.* **199**
52. Modis, Y., and Wierenga, R. K. (1999) A biosynthetic thiolase in complex with a reaction intermediate: the crystal structure provides new insights into the catalytic mechanism. *Structure* **7**, 1279-1290
53. Chun, K. Y., Vinarov, D. A., Zajicek, J., and Miziorko, H. M. (2000) 3-Hydroxy-3-methylglutaryl-CoA synthase. A role for glutamate 95 in general acid/base catalysis of C-C bond formation. *J. Biol. Chem.* **275**, 17946-17953
54. Goblirsch, B. R., Jensen, M. R., Mohamed, F. A., Wackett, L. P., and Wilmot, C. M. (2016) Substrate Trapping in Crystals of the Thiolase OleA Identifies Three Channels That Enable Long Chain Olefin Biosynthesis. *J. Biol. Chem.* **291**, 26698-26706
55. Gajiwala, K. S., Margosiak, S., Lu, J., Cortez, J., Su, Y., Nie, Z., and Appelt, K. (2009) Crystal structures of bacterial FabH suggest a molecular basis for the substrate specificity of the enzyme. *FEBS Lett.* **583**, 2939-2946
56. Aukema, K. G., Kasinkas, L., Aksan, A., and Wackett, L. P. (2014) Use of silica-encapsulated *Pseudomonas* sp. strain NCIB 9816-4 in biodegradation of novel hydrocarbon ring structures found in hydraulic fracturing waters. *Appl. Environ. Microbiol.* **80**, 4968-4976
57. Vararattanavech, A., and Ketterman, A. J. (2007) A functionally conserved basic residue in glutathione transferases interacts with the glycine moiety of glutathione and is pivotal for enzyme catalysis. *Biochem. J.* **406**, 247-256
58. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132
59. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **276**, 307-326
60. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242
61. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255
62. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674
63. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132

64. Omura, S. (1976) The antibiotic cerulenin, a novel tool for biochemistry as an inhibitor of fatty acid synthesis. *Bacteriol. Rev.* **40**, 681-697
65. Ronnett, G. V., Kleman, A. M., Kim, E. K., Landree, L. E., and Tu, Y. (2006) Fatty acid metabolism, the central nervous system, and feeding. *Obesity (Silver Spring)* **14 Suppl 5**, 201S-207S
66. Channon, H. J., and Chibnall, A. C. (1929) The ether-soluble substances of cabbage leaf cytoplasm: The isolation of *n*-nonacosane and di-*n*-tetradecyl ketone. *Biochem. J.* **23**, 168-175
67. Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) The 1.8 A crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. *Structure* **8**, 185-195
68. Jez, J. M., and Noel, J. P. (2000) Mechanism of chalcone synthase. pKa of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase. *J. Biol. Chem.* **275**, 39640-39646
69. Misra, I., and Miziorko, H. M. (1996) Evidence for the interaction of avian 3-hydroxy-3-methylglutaryl-CoA synthase histidine 264 with acetoacetyl-CoA. *Biochemistry* **35**, 9610-9616
70. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345
71. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355-367
72. Garman, E. F., and Owen, R. L. (2006) Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 32-47
73. Burmeister, W. P. (2000) Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 328-341
74. Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J., and Sussman, J. L. (2000) Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 623-628
75. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998) Crystal structure of β-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J.* **17**, 1183-1191
76. Moche, M., Dehesh, K., Edwards, P., and Lindqvist, Y. (2001) The crystal structure of β-ketoacyl-acyl carrier protein synthase II from *Synechocystis* sp. at 1.54 A resolution and its relationship to other condensing enzymes. *J. Mol. Biol.* **305**, 491-503
77. Mathieu, M., Modis, Y., Zeelen, J. P., Engel, C. K., Abagyan, R. A., Ahlberg, A., Rasmussen, B., Lamzin, V. S., Kunau, W. H., and Wierenga,

R. K. (1997) The 1.8 A crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: implications for substrate binding and reaction mechanism. *J. Mol. Biol.* **273**, 714-728
78. Qiu, X., Janson, C. A., Smith, W. W., Head, M., Lonsdale, J., and Konstantinidis, A. K. (2001) Refined structures of β-ketoacyl-acyl carrier protein synthase III. *J. Mol. Biol.* **307**, 341-356
79. Price, A. C., Rock, C. O., and White, S. W. (2003) The 1.3-Angstrom-resolution crystal structure of β-ketoacyl-acyl carrier protein synthase II from *Streptococcus pneumoniae*. *J. Bacteriol.* **185**, 4136-4143
80. Lewis, S. D., Johnson, F. A., and Shafer, J. A. (1976) Potentiometric determination of ionizations at the active site of papain. *Biochemistry* **15**, 5009-5017
81. Qiu, X. Y., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khandekar, S., Lonsdale, J., and Abdel-Meguid, S. S. (1999) Crystal structure of β-ketoacyl-acyl carrier protein synthase III - A key condensing enzyme in bacterial fatty acid biosynthesis. *J. Biol. Chem.* **274**, 36465-36471
82. Bai, T., Zhang, D., Lin, S., Long, Q., Wang, Y., Ou, H., Kang, Q., Deng, Z., Liu, W., and Tao, M. (2014) Operon for biosynthesis of lipstatin, the β-lactone inhibitor of human pancreatic lipase. *Appl. Environ. Microbiol.* **80**, 7473-7483
83. Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2003) Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed. Engl.* **42**, 355-357
84. Lee, M., Gwak, H. S., Park, B. D., and Lee, S.-T. (2005) Synthesis of mycolic acid biosurfactants and their physical and surface-active properties. *J. Am. Oil Chem. Soc.* **82**, 181-188
85. Masamune, S., Hayase, Y., Chan, W. K., and Sobzak, R. L. (1976) Tylonolide hemiacetal, the aglycone of tylosin, and its partial synthesis. *J. Am. Chem. Soc.* **98**, 7874-7875
86. Wyatt, M. A., Ahilan, Y., Argyropoulos, P., Boddy, C. N., Magarvey, N. A., and Harrison, P. H. (2013) Biosynthesis of ebelactone A: isotopic tracer, advanced precursor and genetic studies reveal a thioesterase-independent cyclization to give a polyketide β-lactone. *J. Antibiot. (Tokyo)* **66**, 421-430
87. Wang, Y., Tennyson, R. L., Romo, D. (2004) β-Lactones as intermediates for natural product total synthesis and new transformations. *Heterocycles* **64**, 605-658
88. Albro, P. W., and Dittmer, J. C. (1969) The biochemistry of long-chain, nonisoprenoid hydrocarbons. I. Characterization of the hydrocarbons of *Sarcina lutea* and the isolation of possible intermediates of biosynthesis. *Biochemistry* **8**, 394-404

89. Noyce, D. S., Banitt, E. H. (1966) The stereochemistry of the decarboxylation of β-lactones to form olefins. *J. Org. Chem.* **31**, 4043-4047
90. Mulzer, J., Zippel, M., and Brüntrup, G. (1980) Thermische Decarboxylierung von β-Lactonen: Sterische Mesomeriehinderung als Hinweis auf eine zwitterionische Zwischenstufe. *Angew. Chem.* **92**, 469-470
91. Conti, E., Franks, N. P., and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **4**, 287-298
92. Dick, L. R., Cruikshank, A. A., Grenier, L., Melandri, F. D., Nunes, S. L., and Stein, R. L. (1996) Mechanistic studies on the inactivation of the proteasome by lactacystin: a central role for clasto-lactacystin β-lactone. *J. Biol. Chem.* **271**, 7273-7276
93. Rachid, S., Huo, L., Herrmann, J., Stadler, M., Kopcke, B., Bitzer, J., and Muller, R. (2011) Mining the cinnabaramide biosynthetic pathway to generate novel proteasome inhibitors. *ChemBioChem* **12**, 922-931
94. Zhao, C., Coughlin, J. M., Ju, J., Zhu, D., Wendt-Pienkowski, E., Zhou, X., Wang, Z., Shen, B., and Deng, Z. (2010) Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an "acyltransferase-less" type I polyketide synthase that incorporates two distinct extender units. *J. Biol. Chem.* **285**, 20097-20108
95. Hemmerling, F., and Hahn, F. (2016) Biosynthesis of oxygen and nitrogen-containing heterocycles in polyketides. *Beilstein J. Org. Chem.* **12**, 1512-1550
96. Nichols, D. S., Nichols, P. D., and McMeekin, T. A. (1995) A new n-C31:9 polyene hydrocarbon from Antarctic bacteria. *FEMS Microbiol. Lett.* **125**, 281-285
97. Blom, T., Somerharju, P., and Ikonen, E. (2011) Synthesis and biosynthetic trafficking of membrane lipids. *Cold Spring Harb. Perspect. Biol.* **3**, a004713
98. Pawar, S., and Schulz, H. (1981) The structure of the multienzyme complex of fatty acid oxidation from *Escherichia coli*. *J. Biol. Chem.* **256**, 3894-3899
99. Haase, I., Grawert, T., Illarionov, B., Bacher, A., and Fischer, M. (2014) Recent advances in riboflavin biosynthesis. *Methods Mol. Biol.* **1146**, 15-40
100. Ladenstein, R., Fischer, M., and Bacher, A. (2013) The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system. *FEBS J.* **280**, 2537-2563
101. Smith, J. L., Skiniotis, G., and Sherman, D. H. (2015) Architecture of the polyketide synthase module: surprises from electron cryo-microscopy. *Curr. Opin. Struct. Biol.* **31**, 9-19

102. Robbins, T., Liu, Y.-C., Cane, D. E., and Khosla, C. (2016) Structure and mechanism of assembly line polyketide synthases. *Curr. Opin. Struct. Biol.* **41**, 10-18
103. Enderle, M., McCarthy, A., Paithankar, K. S., and Grininger, M. (2015) Crystallization and X-ray diffraction studies of a complete bacterial fatty-acid synthase type I. *Acta Crystallographica Section F* **71**, 1401-1407
104. Boehringer, D., Ban, N., and Leibundgut, M. (2013) 7.5-A cryo-em structure of the mycobacterial fatty acid synthase. *J. Mol. Biol.* **425**, 841-849
105. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254
106. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850-858
107. Nadano, D., Aoki, C., Yoshinaka, T., Irie, S., and Sato, T. A. (2001) Electrophoretic characterization of ribosomal subunits and proteins in apoptosis: specific downregulation of S11 in staurosporine-treated human breast carcinoma cells. *Biochemistry* **40**, 15184-15193
108. Galbraith, C. G., and Galbraith, J. A. (2011) Super-resolution microscopy at a glance. *J. Cell Sci.* **124**, 1607-1611